No Mere Deodands: Human Responsibilities in the Use of Violent Intelligent Systems Under Public International Law

No Mere Deodands:

Human Responsibilities in the Use of Violent Intelligent Systems under Public International Law

Lucas V.M. Bento

A Thesis

in the Field of Government

in Partial Fulfillment of the Requirements for

the Master of Liberal Arts Degree

Harvard University

March 2017

Abstract

The tide toward the militarization of autonomous technologies has prompted critics to propose a pre-emptive ban on their development for fear that they may not adhere to international laws and, worse still, that no one will be responsible for their use. These criticisms, however, are rooted in pessimistic prognoses that misconstrue the potential of emerging technologies and international law's ability to regulate them. Accordingly, this thesis advances three arguments to dismantle these dystopian perspectives. First, a pre-emptive ban ignores the centuries-long distribution of violent tasks between humans and non-human actants. In the process, it seeks to revise current terminologies by shifting the focus on autonomy and lethality toward intelligence, violence, and systems. Second, the current international legal architecture is adequate to (a) ensure the responsible use of emerging autonomous weapons systems and (b) allocate human responsibility for their use. Critics often argue that it is impossible to preprogram all eventualities of warfare into a machine, but this perspective ignores advances in machine learning, that enable intelligent systems to teach themselves rules based on set parameters and algorithms. Third, critics misunderstand the networked nature of human violence, and consequently underestimate the elasticity of international law. To this end, this thesis borrows from evolutionary biology, psychology, and semiotics to explain the composition and constitution of networks of human violence. Ultimately, by viewing armed conflicts through the lenses of networks, this thesis argues that international law is capable of regulating human violence regardless of its conduit.

To Margarida, Valdir, Thiago, Lauren and Theodor for their love and support,

And to all other members of our family—

Including those that have carried the burden and are now at rest,

And those yet unborn for whom we must fulfil the same responsibility, and

Only rest when we leave behind a better world than the one we inherited.

"We can be humble and live a good life with the aid of machines, or we can be arrogant and die."

Norbert Wiener

Acknowledgments

I would like to thank my Thesis Director, Professor Gabriella Blum for her inspiration and generous guidance throughout my research. I am extremely fortunate to have benefited from her visionary expertise and I will remain forever indebted to her time, direction and feedback. I would like to thank my Research Advisor, Dr. Donald Ostrowski, for his comments on my proposal and earlier drafts, as well as Chuck Houston and Stephen Blinn. I am also grateful to the University for its leadership in and commitment to widening access to knowledge. As this thesis was written during stolen family time, I would like to thank my family for their love, patience, and support. I am grateful to my parents for their encouragement. To all roboticists out there seeking to develop the best iteration of a man, look no further than my brother Thiago for a blueprint; his integrity, character, and neverending curiosity form the basis of everything that I strive to become. To my wife Lauren: despite my optimism about artificial intelligence, I am certain that no machine learning will ever be able to match the distinction of your soul. The five hundred footnotes lurking in the details collected here represent an almost equal number of nights where I shared my time and presence with the ideas assembled in this draft. This work would not have been possible without your patience and support. And to the beautiful algorithms we created, Theo et al.: thank you for energizing my search for the Truth. It is more important to emphasize that the above should not be taken to agree with any of the views herein, nor with my expression of them. All errors, big or small, remain my own.

Table of Contents

List of Figures

Chapter I

Introduction

To study war is to study the expressions of human power through violence which, despite diverse materializations, remain products of human will. While human violence has historically been declining,[1] our species' adeptness at innovating how violence is produced has dramatically increased.[2] From the arrow to the algorithm, means and methods of human violence have evolved to enable humans to do more with less human input.[3] The study of war is thus also a study of how humans produce, organize, and engineer violence through "bureaucratic death-dealing enterprise[s]"[4] that combine human actors and non-human actants. This re-materialization of violence—not to be confused with its de-humanization—entails significant legal consequences because it challenges the anthropomorphic conception of the legal subject and how legal systems attribute responsibility.

---

[1] STEVEN PINKER, THE BETTER ANGELS OF OUR NATURE xxi (Penguin Books, 2011) ("[V]iolence has declined over long stretches of time, and today we may be living in the most peaceable era in our species' exsitence.").

[2] BENJAMIN WITTES AND GABRIELLA BLUM, THE FUTURE OF VIOLENCE 7 (Basic Books, 2015) ("Technologies that put destructive power traditionally confined to states in the hands of small groups and individuals have proliferated remarkably far, as a general matter. That proliferation is accelerating at an awe-inspiring clip across a number of technological platforms—in particular, networked computers and biotechnology and, in the not-so-distant future, robotics and nanotechnology as well…They are platform technologies—that is, technologies that facilitate generative creativity in their users to build and invent new things, new weapons, and new modes of attack.").

[3] PINKER, *supra* note 1, at 673 ("Over the millennia weapons, just like every technology, got better and better").

[4] ROSA BROOKS, HOW EVERYTHING BECAME WAR AND THE MILITARY BECAME EVERYTHING 4 (Simon & Schuster, 2016).

These issues have come to the fore in the context of weapon systems capable of autonomously participating in—or actually carrying out—the infliction of lethal force in international armed conflicts.[5] While these technologies already exist in some form, current apprehensions concern the development of what some have dubbed "Lethal Autonomous Robots" ("LARs") or "Lethal Autonomous Weapon Systems" ("LAWS")—weapon systems capable of selecting, targeting, and firing at a target without human intervention. Numerous countries are studying the capabilities of these systems, and some, like the US, Russia, and China, have expressed an interest to utilize them at some point in the near future.

The tide toward the militarization of autonomous technologies has prompted a handful of governments,[6] organizations, and advocates to propose a "ban" or "moratorium" on the development of such systems for fear that they may not adhere to international laws and, worse still, that no one will be responsible for their use. This potential accountability vacuum,[7] or responsibility gap,[8] effectively elevates

---

[5] *See e.g.* Nils Melzer, *Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare*, European Parliament, Directorate-General for External Policies, 43 (May 2013), *available at* http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/410220/EXPO-DROI_ET(2013)410220_EN.pdf ("How does increasing robotic autonomy affect the legal responsibility of States and individuals for potential harm which may result from their use?").

[6] *See e.g.* Austria, Chile, Costa Rica, Ecuador, Germany, Mexico, Pakistan, Poland, and Zambia. *See* Dustin A. Lewis, Gabriella Blum, and Naz K. Modirzadeh, *War-Algorithm Accountability* 62, (Harvard Law School Program on International Law and Armed Conflict, 2016), *available at* http://pilac.law.harvard.edu/waa/

[7] Christof Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, ¶ 77, Human Rights Council, 3rd Session, A/HRC/23/47 (Apr. 9, 2013).

[8] *See e.g.* Matthias, A., *The responsibility gap: Ascribing responsibility for the actions of learning automata*, 6(3) ETHICS AND INFORMATION TECHNOLOGY, 175–183 (2004); Heyns, *supra* 7, ¶ 77.

autonomous weapons to "deodands"—a peculiar class of moving objects, like a rock,[9] or "beasts,"[10] like a horse, whose "movement," despite being the "immediate occasion of a fatal accident,"[11] entails no human liability. Developed in the 11th century to make legal sense of deaths or injuries caused by "moveable thing[s] inanimate, or beast[s] animate,"[12] deodand law is a reflection of jurisprudential struggles to attribute responsibility for human fatalities caused by non-human actants.[13]

While the critics' concerns about the dangers of autonomous weapons are understandable, they are rooted on pessimistic prognoses that misconstrue the potential of emerging technologies[14] and international law's ability to regulate them. Accordingly, this thesis advances three arguments to dismantle these dystopian perspectives. The thesis' structure tracks these arguments in sequential order.

First, advocates for a pre-emptive ban assume that autonomous technologies are unique because they shift powers of lethal decision-making away from human

---

[9] Anna Pervukhin, *Deodands: A Study in the Creation of Common Law Rules*, 47(3) THE AMERICAN JOURNAL OF LEGAL HISTORY, 237, 238 ("If a horse kicked a man in the chest or a stone fell on his head, the horse or the stone would be a deodand.").

[10] *Id.*, 238.

[11] *Id.* ("an animal or object was a deodand if its movement was the immediate occasion of a fatal accident.").

[12] SIR EDWARD COKE, INSTITUTES OF THE LAWS OF ENGLAND, VOLUME III 58 (Printed for A. Crooke [and 12 others], 1669) ("when any moveable thing inanimate, or beast animate, does move to, or cause the untimely death of any reasonable creature by mischance…without the will, offence, or fault of himself, or of any person, [that thing is a deodand.]"); Pervukhin, *supra* note 9, 238; *see also* Austin v. United States, 509 U.S. 602, 628 (1993)("[I]f a man was killed by a moving cart, the cart and its horses were deodands…").

[13] The law of deodands was abolished by The Deodands Act 1846 (9 & 10 Vict, c.62).

[14] Deborah G. Johnson, *Technology with No Human Responsibility?*, 127(4) JOURNAL OF BUSINESS ETHICS 707, 709 (2014) ("speculations about a responsibility gap misrepresent the situation and are based on false assumptions about technological development and about responsibility.").

control. This assumption, however, ignores the centuries-long distribution of violent tasks between humans and non-human actants. This thesis employs the term "actant" as a "semiotic definition . . . that is something that acts or to which activity is granted by another . . . an actant can literally be anything provided it is granted to be the source of action."[15] Practical responsibilities for the production of violence (from planning to infliction) are inextricably shared between humans and other actants. From the automated assembly lines that manufacture weapon parts, through the computer systems that guide fire and forget missiles, to the warhead that inflicts lethal force, the *praxis* of violence is not confined to the human hand—it is a product of distributed practical responsibilities in networks of violence. This distribution is at the heart of how humans design violence in international armed conflicts. While human beings have always been functionally and legally responsible for organizing violence, technology has allowed our species to distribute numerous tasks to non-human actants, including assembly lines, tanks, airplanes, weapons, drones, and, more recently, artificial intelligence.

Terminology is also making it harder to engage in meaningful debates about the legal and moral implications of autonomous weapons. In particular, the focus on "autonomy," when combined with other terms like "lethality" and "robot," distracts stakeholders from International Humanitarian Law ("IHL")'s[16] current priority to maximize human security through the moderation of the means and methods of

---

[15] Bruno Latour, *On actor network theory: a few clarifications*, 47 SOZIALE WELT, 360, 373 (1996).

[16] TERRI THORNE AND ERIC C. HUSBY, LAW OF ARMED CONFLICT DESKBOOK 8 (The Judge Advocate General's Legal Center and School, U.S. Army, 2012), *available at* https://www.loc.gov/rr/frd/Military_Law/pdf/LOAC-Deskbook-2012.pdf ("The law of armed conflict is also referred to as the law of war (LOW) or international humanitarian law (IHL)").

warfare. This thesis proposes an alternative taxonomy, namely Violent Intelligent Systems ("VIS"), to facilitate discourse concerning weapons capable of autonomous operation.[17]

Second, current international laws are adequate to ensure the responsible use of the "means and methods" of warfare, including emerging autonomous weapons. The well-established principles of distinction, proportionality, precautionary measures, necessity, and doubt under IHL are focused on moderating the impact of violence regardless of its conduit. Critics often argue that it is impossible to preprogram all eventualities of warfare into a machine, and accordingly, it is impossible, for instance, to teach a machine how to distinguish civilians from combatants, or to gauge the value of military targets. But this perspective ignores advances in artificial intelligence, particularly machine learning, that enable intelligent systems to teach themselves rules based on set parameters and algorithms. If the Industrial Revolution automated manual tasks and the Information Age automated mental tasks, "the 'Machine Learning' Revolution is automat[ing] automation itself."[18] Since calls for a preventive ban are based on a thought experiment that focuses on technological incompetence, this thesis, in seeking to develop a more optimistic narrative, engages in a thought experiment about technology's potential based on current developments. As Johnson put it, "[t]he

[17] *See also* Lewis, Blum, and Modirzadeh, *supra* note 6, at vii ("the [Autonomous Weapon Systems] framing has largely precluded meaningful analysis of whether it (whatever "it" entails) can be regulated, let alone whether and how it should be regulated….[and proposing] the concept of war algorithm [,which includes] any algorithm that is expressed in computer code, that is effectuated through a constructed system, and that is capable of operating in relation to armed conflict. Those algorithms seem to be a—and perhaps the—key ingredient of what most people and states discuss when they address AWS.").

[18] PEDRO DOMINGOS, THE MASTER ALGORITHM 10 (Basic Books, 2015).

discourse on responsibility and artificial agents . . . is largely a discourse about the future."[19] This thesis argues that international responsibility for the use of VIS falls squarely on humans and human institutions, e.g., States, as dictated by principles of State responsibility and international criminal law.[20]

Third, critics misunderstand how human violence is organized in international armed conflicts, and consequently underestimate the elasticity of international law. Human violence does not operate in a vacuum. It is a product of networks that combine humans and non-human actants to deliver force on an adversary. In these networks, legal responsibilities do not travel with practical responsibilities,[21] but firmly remain with the designers and manufacturers of violence: humans and human institutions. As Gabriella Blum and Natalie Lockwood put it, "wars do not merely 'happen'; they are waged and prosecuted. Human beings control the incidence of armed conflicts and what transpires in their course."[22] To this end, this thesis borrows from evolutionary biology, psychology, and semiotics to explain the composition and constitution of networks of human violence. It argues that violent technologies, including Violent Intelligent Systems, are phenotypic expressions of the human drive to wield political power through violence. In this vein, VIS are active

---

[19] Johnson, *supra* 14, 710.

[20] Heyns, *supra* 7, at 15 ("The question of legal responsibility could be an overriding issue.").

[21] *See also* H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 212-214 (Oxford University Press, 2nd edition, 2008) (distinguishing between role responsibility and cause responsibility).

[22] Gabriella Blum and Natalie Lockwood, *Earthquakes and Wars: the Logic of International Reparations*, in MAY, L. AND EDENBERG, E., (eds.), "JUS POST BELLUM" AND TRANSITIONAL JUSTICE 190 (Cambridge University Press, 2013).

external media that assist the cognitive organization of violence much like a pen and calculator helped organize information processing in the Information Age. Code, as the weapon of the 21^{st} century, acts as a cognitive agent of human intentionality. Ultimately, by viewing armed conflicts through the lenses of networks, this thesis argues that international law is capable of regulating means and methods of *human violence* (i.e., violence designed by humans to be unleashed in many ways) and not just *human* means and methods of violence (i.e., ways in which violence is unleashed only by humans).

Chapter II

Human Design

Violent Intelligent Systems in the 21st Century

Violent Intelligent Systems ("VIS") increasingly capable of autonomous operation are no longer the product of the human imagination. These technologies are no longer fanciful mockups of archetypal Renaissance men, such as Leonardo da Vinci's "mechanical knight,"[23] or fictional characters, such as Lord Dunsany's "thinking beast of steel"[24] or James Cameron's "The Terminator." Weapon systems capable of autonomous action have now entered the realm of science fact with real legal, military and geopolitical implications.[25]

Indeed, weapons with varying degrees of autonomy are already in use. Drones, as unmanned weapons, are widely considered precursors to autonomous weapons and have become staple military tools used in various capacities by over 85 armed forces around the world.[26] Highly sophisticated weapons are already playing autonomous defensive roles.[27] For example, some U.S. Navy ships are equipped with

---

[23] M.E. Moran, *The da Vinci robot*, 20 JOURNAL OF ENDOUROLOGY 12, 986–990 (2006).

[24] LORD DUNSANY, THE LAST REVOLUTION 1 (Talos, reprint edition, 2015).

[25] Benjamin Kastan, *Autonomous Weapons Systems: A Coming Legal 'Singularity'?*, 1 JOURNAL OF LAW, TECHNOLOGY, & POLICY 1, 45 (2013) ("In recent years, this technology has moved from the realm of science fiction to reality").

[26] *World of Drones: Military*, International Security Website, *available at* http://securitydata.newamerica.net/world-drones.html [Last accessed September 26, 2016].

[27] Autonomous weapons with offensive capabilities are yet to be deployed. This is unsurprising because offensive autonomy arguably raises more complex issues since their use

the Phalanx anti-ship missile system "capable of autonomously performing its own search, detect, evaluation, track, engage and kill assessment functions."[28] South Korea's SGR-A1 sentry guards the demilitarized zone separating North and South Korea, with capabilities to autonomously identify and destroy targets.[29] Israel's Harpy anti-radar missile system and the Guardium sentry robot are capable of autonomously destroying enemy radars and identifying suspicious elements on the border with Gaza.[30] Russia has reportedly installed robot sentries at ballistic missile installations to "detect and destroy targets without a human giving the go-ahead."[31]

In both policy and practical terms, systems capable of autonomous operation are gradually becoming central to the military strategy of some of the world's largest militaries, including the United States. In the last two decades, US policy has gradually shifted from a focus on developing unmanned vehicles, such as drones, to increasing autonomous capabilities. In 2001, the U.S. Congress issued a mandate stating that by 2010 one-third of all "U.S. deep-strike aircraft should be unmanned"

---

increases the propensities of engagement due to increased mobility and greater exposure to the enemy and noncombatants.

[28] *MK 15-Phalanx Close-In Weapons System,* U.S. Navy Website, *available at* http://www.public.navy.mil/surfor/Pages/Phalanx-CIWS.aspx [Last accessed September 26, 2016].

[29] Alexander Velez-Green, *The South Korean Sentry- A "Killer Robot" to Prevent War*, LAWFARE BLOG, March 1, 2015, *available at* http://www.lawfareblog.com/2015/03/the-foreign-policy-essay-the-south-korean-sentry-a-killer-robot-to-prevent-war/.

[30] Pablo Kalmanovitz, *Judgment, liability, and the risk of riskless warfare*, in NEHAL BHUTA, SUSANNE BECK, ROBIN GEISS, HIN-YAN LIU AND CLAUS KRESS, AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, AND POLICY 147 (Cambridge University Press, 2016).

[31] Randy Rieland, *Can Killer Robots Learn to Follow the Rules of War?*, SMITHSONIAN MAGAZINE, May 29, 2014, *available at* http://www.smithsonianmag.com/innovation/can-killer-robots-learn-follow-rules-war-180951581/?no-ist [Last accessed September 26, 2016].

and that by 2015 "one-third of all ground vehicles should likewise be unmanned."[32] While that goal was not met, in 2009 the Department of Defense ("DoD") emphasized that "the level of autonomy [in unmanned systems] should continue to progress from [a] fairly high level of human control/intervention to a high level of autonomous tactical behavior that enables more timely and informed human oversight."[33] Two years later, in its 2011-2036 Roadmap, the DoD stated that it must "continue to pursue technologies and policies that introduce a higher degree of autonomy to reduce the manpower burden and reliance on full-time high-speed communications links while also reducing decision loop cycle time."[34]

In 2012, the DoD released Directive 3000.09 ("Directive"), entitled "Autonomy in weapon systems," which aims to:

> a. Establish DoD policy and assign responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems, including manned and unmanned platforms.
>
> b. Establish guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements.[35]

---

[32] Gary E. Marchant, Braden Allenby, Ronald Arkin, Edward T. Barrett, Jason Borenstein, Lyn M. Gaudet, Orde Kittrie, Patrick Lin, George R. Lucas, Richard O'Meara, Jared Silberman, *International Governance of Autonomous Military Robots*, THE COLUMBIA SCIENCE AND TECHNOLOGY LAW REVIEW, Vol. XII, 277 (2011).

[33] *FY2009-2034 Unmanned Systems Integrated Roadmap*, 27, U.S. DEPARTMENT OF DEFENSE *available at* http://www.acq.osd.mil/sts/docs/UMSIntegratedRoadmap2009.pdf .

[34] *FY 2011-2036 Unmanned Systems Integrated Roadmap*, U.S. DEPARTMENT OF DEFENSE, vi, *available at* http://www.acq.osd.mil/sts/docs/Unmanned%20Systems%20Integrated%20Roadmap%20FY2011-2036.pdf.

[35] Directive Number 3000.09, U.S. Department of Defense, November 21, 2012, *available at* http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf [hereinafter "Directive Number 3000.09"] at ¶ 1.

The Directive provides that "[a]utonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to *exercise appropriate levels of human judgment* over the use of force."[36] It is unclear, however, what metric or standard is used to determine the appropriateness of "human judgment," which is also an undefined term. One interpretation is to define it against what is permissible under the Directive. First, the Directive provides that semi-autonomous systems may be used to apply lethal or non-lethal force. Second, human-supervised autonomous weapon systems may be used to select and engage targets, but not human targets, for local defense purposes. These guidelines mean the DoD will not, as a matter of standard policy,[37] use autonomous weapons systems to inflict lethal force. Third, autonomous weapons systems may be used to apply non-lethal, non-kinetic force, such as electronic attacks, against material targets. While the use of lethal autonomous force against human targets is presumptively forbidden, the guidelines provide that use that "fall[s] outside" standard policy is allowed as long as it is approved by the Under Secretary of Defense for Policy, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the [Chairman of the Joint Chiefs of Staff], "before formal development and again before fielding."[38] These approvals add a layer of human judgment to the development and deployment of

---

[36] *Id.*, at ¶ 4.(a) (emphasis added).

[37] U.S. Opening Statement at the CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems, April 13, 2015, *available at* https://geneva.usmission.gov/2015/04/15/u-s-opening-statement-at-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapons-systems/ ("The US delegation to the 2015 CCW meeting made "clear that the Directive does not establish a U.S. position on the potential future development of lethal autonomous weapon systems – it neither encourages nor prohibits the development of such future systems.").

[38] Directive Number 3000.09, *supra* note 35, at ¶ 4(d).

lethal autonomous weapon systems. It suggests that high-level approvals will be necessary for mission-specific situations where lethal force is used, or against a specific set of combatants or enemies.[39]

The Directive must be understood in the context of the Defense Innovation Initiative, also known as the Pentagon's Third Offset Strategies,[40] designed to inter alia "accelerate innovation ... [to] identify, develop, and field breakthrough technologies and systems that sustain and advance the capability of U.S. military power."[41] The First Offset Strategy, officially called the New Look Program and steered by President Einsenhower, sought to use nuclear weapons as a method of deterrence to offset Soviet power after the Second World War. The Second Offset Strategy was the development of precision munitions and stealth capabilities to offset enemy air and ground forces, which the US could not, at the time, match. As the world enters the "most volatile security environments … in decades,"[42] the Third

[39] The directive's attempt to "avoid organized irresponsibility" has been criticized because "it does not clearly address a fundamental aspect of fully autonomous systems—namely, that a system's course of action is not necessarily completely predictable for the operator." Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems*, 47 VANDERBILT JOURNAL OF TRANSNATIONAL LAW, 1, 33 (2014).

[40] Bob Work, *The Third Offset Strategy and its Implications for Partners and Allies*, Deputy Secretary of Defense Speech, U.S. Department of Defense (Jan. 28, 2015) *available at* http://www.defense.gov/News/Speeches/Speech-View/Article/606641/the-third-us-offset-strategy-and-its-implications-for-partners-and-allies ("[T]hat is the department's innovation initiative, the Defense Innovation Initiative, and what we refer to right now as the third offset strategy, or perhaps more accurately to everyone here, offset strategies.").

[41] Chuck Hagel, *Defense Innovation Initiative Memorandum*, U.S. DEPARTMENT OF DEFENSE, 2 (Nov. 15, 2014) *available at* http://www.defense.gov/Portals/1/Documents/pubs/OSD013411-14.pdf.

[42] Bob Work, *supra* note 40.

Offset Strategies seek to "identify and invest in innovative ways to sustain and advance America's military dominance for the 21st century."[43]

These offset strategies recognize that the distribution of violence between humans and non-human actants (e.g., nuclear weapons, precision munitions) are critical to achieve a competitive advantage on the battlefield. Indeed, President Eisenhower's First Offset Strategy was designed to reduce manpower while maintaining deterrence.[44] The Second Offset Strategy sought to continue this trajectory by employing more precise and less destructive technologies. A critical component of the Third Offset Strategies, however, is that it focuses on human-technology collaboration. As the US Deputy Secretary of Defense put it, "the big idea right now for deterrence is human-machine collaboration and combat teaming."[45] The idea that autonomous weapons will simply be deployed and left to their own devices is not what is envisaged here. As this thesis argues, human violence in international armed conflicts has always been expressed as a network where human and non-human actants collaborate in some form, rather than autonomous lethal systems that operate independently from human control.

---

[43] *Id.*

[44] *Id.* ("So to counter Soviet superiority without bankrupting the West, Eisenhower directed a top-level strategic review which resulted in what was called the New Look. And that said the U.S. would reduce military manpower and would rely instead on its nuclear arsenal, where we had a big advantage at the time, for deterrence."); *see also generally* Geoffrey Parker, *Introduction: The Western Way of War*, in GEOFFREY PARKER (ed.), THE CAMBRIDGE HISTORY OF WARFARE 1 (Cambridge University Press, 2005) ("The western way of war, which also boasts great antiquity, rests upon five principal foundations. First, the armed forces of the West have always placed heavy reliance on superior technology, usually to compensate for their inferior numbers.").

[45] Cheryl Pellerin, *Work: Human-Machine Teaming Represents Defense Technology Future*, U.S. DEPARTMENT OF DEFENSE NEWS, November 8, 2015, *available at* http://www.defense.gov/News/Article/Article/628154/work-human-machine-teaming-represents-defense-technology-future.

To achieve these strategies, the DoD is investing in numerous programs to better understand and develop autonomous capabilities.[46] For example, the Probabilistic Programming for Advancing Machine Learning seeks to "create an advanced computer programming capability that greatly facilitates the construction of new machine learning applications in a wide range of domains," including "autonomous system navigation and control."[47] The Cognitive Computing Project focuses on developing "technologies that enable[] computing and autonomy systems to learn and apply knowledge gained through experience,"[48] including "systems with increased self-reliance and the capacity to operate with reduced programmer and operator intervention."[49] The Autonomous Robotic Manipulation program "developed advanced robotic technologies that enabled autonomous (unmanned) mobile platforms to manipulate objects without human control or intervention."[50] The Pentagon is also "testing autonomous ships that can remain at sea for months

---

[46] The DoD also opened an "outreach center" in Silicon Valley to "encourage some of the most innovative start-ups to turn their attention to national security." *See* Christian Davenport, *Robots, swarming drones and 'Iron Man': Welcome to the new arms* race, THE WASHINGTON POST (Jun. 17, 20160 *available at* https://www.washingtonpost.com/news/checkpoint/wp/2016/06/17/robots-swarming-drones-and-iron-man-welcome-to-the-new-arms-race/.

[47] *Fiscal Year (FY) 2017 President's Budget Submission*, U.S. DEPARTMENT OF DEFENSE, *available at* http://www.darpa.mil/attachments/DARPAFY17PresidentsBudgetRequest.pdf at 14-15.

[48] *Fiscal Year (FY) 2016 President's Budget Submission*, U.S. DEPARTMENT OF DEFENSE, *available at* http://www.darpa.mil/attachments/(2G1)%20Global%20Nav%20-%20About%20Us%20-%20Budget%20-%20Budget%20Entries%20-%20FY2016%20(Approved).pdf [hereinafter DARPA 2016 Budget] at 107.

[49] DARPA 2016 Budget, *supra* note 48, at 107.

[50] DARPA 2016 Budget, *supra* note 48, at 109 ("A key objective was intelligent control of manipulators to independently perform subtasks over a broad range of domains of interest to the warfighter, thereby reducing operator workload, time on target, training time, bandwidth, and hardware complexity.").

without a crew."[51] In 2014, the Office of Naval Research awarded $7.5 million in grant money to university researchers to explore how to, inter alia, develop "novel computational means by which [autonomous] robots can reason and act ethically in the face of complex, practical challenges."[52] The US and UK armed forces are also investing in joint programs to "deepen" collaboration in the development of autonomous systems.[53] These investments have led some to prophesize that "[o]perational realities will likely drive the United States to discard its practice of keeping a human in the loop for lethal targeting decisions."[54]

Other militaries around the world are following suit. For example, by 2025, "30 percent of all military technology in the Russian Armed Forces is expected to consist of robotic hardware."[55] China is similarly investing in military robotics and autonomy, having unveiled robots capable of "wield[ing] anti-tank weapons, grenade launchers and assault rifles."[56] The Dutch Armed Forces also noted that if they "are

---

[51] Davenport, *supra* note 46.

[52] *Moral Competence in Computational Architectures for Robots*, Human Robot Interaction Laboratory, (Apr. 8, 2016), *available at* http://hrilab.tufts.edu/muri13/

[53] Philip Dunne, *UK and US look to robotics help for 'last mile',* U.K. MINISTRY OF DEFENSE NEWS (Jul. 14, 2016), *available at* //www.gov.uk/government/news/uk-and-us-look-to-robotics-help-for-last-mile

[54] Michael N. Schmitt and Jeffrey S. Thurnher, *"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231, 237 (2013).

[55] Franz-Stefan Gady, *Meet Russia's New Killer Robot*, THE DIPLOMAT (Jul. 21, 2015), *available at* http://thediplomat.com/2015/07/meet-russias-new-killer-robot/.

[56] Patrick Tucker, *The Pentagon is Nervous about Russian and Chinese Killer* Robots, DEFENSE ONE (Dec. 14, 2015), *available at* http://www.defenseone.com/threats/2015/12/pentagon-nervous-about-russian-and-chinese-killer-robots/124465/?oref=d-river

to remain technologically advanced, autonomous weapons will have a role to play, now and in the future."[57]

Autonomous weapons will likely first be used for specific missions that do not employ lethal force, and gradually progress into—or "incremental[ly] march"[58] toward—lethal capabilities as technologies become more sophisticated and better integrated with human networks.[59] As more militaries pursue these programs, the research, development and eventual deployment of these technologies will intensify around the world.[60] If we are to take unmanned aerial vehicles as an indicator, the number of countries that acquired an unmanned aerial vehicle (UAV) system nearly doubled in the past decade alone.[61]

---

[57] *Autonomous weapon systems: the need for meaningful human control*, Advisory Report, DUTCH ADVISORY COUNCIL ON INTERNATIONAL AFFAIRS (Mar. 2, 2016) *available at* http://aiv-advies.nl/8gr.

[58] Kenneth Anderson and Matthew Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won 't Work and How the Laws of War Can,* HOOVER INSTITUTION RESEARCH PAPERS, 4-6 (Apr. 9, 2013) *available at* http://www.hoover.org/publications/monographs/144241.

[59] *Unmanned Aircraft Systems Flight Plan 2009-2047*, United States Air Force, 41 (May 18, 2009) *available at* http://www.fas.org/irp/program/collect/uas_2009.pdf (For example, the Flight Plan for the US Air Force notes that "[a]dvances in computing speeds and capacity will change how technology affects the [Observe, Orient, Decide, and Act] loop. As a result, "humans will no longer be 'in the loop' but rather 'on the loop' – monitoring the execution of certain decisions. Simultaneously, advances in [artificial intelligence] will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input.").

[60] BROOKS, *supra* note 4, at 135 ("the fact that other states, including some U.S. adversaries, are pursuing research into such weapons systems means that the United States may have to do the same, for purely defensive purposes: human reaction time won't be able to keep pace with machine reaction time.").

[61] JEFFREY L. CATON, AUTONOMOUS WEAPON SYSTEMS: A BRIEF SURVEY OF DEVELOPMENTAL OPERATIONAL, LEGAL, AND ETHICAL ISSUES 8, The Letort Papers (United States Army War College Press, 2015).

This trajectory is man-made—there is no "blind watchmaker."[62] The development of system autonomy is the product of human-designed endeavors, both at the individual and institutional levels.[63] The systems are not self-reproducing—they are products of research programs that incorporate technologies developed by human policy makers, military personnel, developers, engineers, and others in public and private sectors.

Hostility toward the Technological Distribution of Violence Tasks

The tide toward the militarization of autonomy has raised alarm in government circles and civic society worldwide.[64] Concerned stakeholders have argued that autonomous weapons should be banned because they will be incapable to meet the stringent standards of international humanitarian law,[65] will make war more

---

[62] JOHN MARKOFF, MACHINES OF LOVING GRACE xiv (HarperCollins, 2015) ("There's no blind watchmaker for the evolution of machines. Whether we augment or automate is a design decision that will be made by individual human beings.").

[63] Johnson, *supra* note 14, at 6 ("Given that producing a new technology involves many human actors making decisions and getting others to accept those decisions, in order to imagine a future time at which there will be artificial agents for which no humans are responsible, we have to imagine that the human actors involved would decide to create, release, and accept technologies that are incomprehensible and out of the control of humans.").

[64] *See e.g. Losing Humanity: The Case against Killer Robots*, HUMAN RIGHTS WATCH (Nov. 2012)[hereinafter "Losing Humanity"]; Campaign to Stop Killer Robots, *available at* https://www.stopkillerrobots.org/; Republic of Sierra Leone, Opening Statement by Ambassador Yvette Stevens, CCW 2016 Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), at 3 (Apr. 2016) *available at* http://www.unog.ch/80256EDD006B8954/(httpAssets)/0054AE2FAA24E566C1257F9B004A2CAB/$file/SIERRA+LEONE+GENERAL+STATEMENT+2016+MEETING+ON+LAWS.pdf.

[65] Losing Humanity, *supra* note 64, at 1; *see also* Hin-Yan Liu, *Categorization and legality of autonomous and remote weapons systems*, 94 INTERNATIONAL REVIEW OF THE RED CROSS, No. 886, 630 (2012) ("International humanitarian law (IHL) in its current manifestation is insufficient to regulate the growing use of autonomous and remote weapons systems . . ."); Melzer, *supra* note 5, at 1

likely,[66] and will effectively allow humans to evade responsibility.[67] From a moral standpoint, some have also argued that such weapons should be banned because decisions of life and death involve moral judgments that only humans can make.[68] Ultimately, these concerns focus on notions of responsibility that shape our understanding of permissible conduct: the ability to act responsibly and to be responsible for one's acts.[69]

The drive to ban the development of autonomous weapon systems have been largely set in motion by non-government organizations, such as Human Rights Watch ("HRW") and the Campaign to Stop Killer Robots. In *Losing Humanity,* HRW has

---

("the EU should work towards the adoption of a binding international agreement, or a non-binding code of conduct, aiming to restrict the development, proliferation or use of certain unmanned weapon systems in line with the legal consensus achieved.").

[66] *Mind the Gap: The Lack of Accountability for Killer Robots*, HUMAN RIGHTS WATCH (2015) [hereinafter "Mind the Gap"], at 1 ("potential threats include the prospect of an arms race and proliferation to armed forces with little regard for the law. "); Heyns, *supra* note 7, at 11 ("Due to the low or lowered human costs of armed conflict to States with LARs in their arsenals, the national public may over time become increasingly disengaged and leave the decision to use force as a largely financial or diplomatic question for the State, leading to the "normalization" of armed conflict. LARs may thus lower the threshold for States for going to war or otherwise using lethal force, resulting in armed conflict no longer being a measure of last resort ").

[67] Mind the Gap, *supra* note 66, at 2 ("Existing mechanisms for legal accountability are ill suited and inadequate to address the unlawful harms fully autonomous weapons might cause.").

[68] *See* Campaign to Stop Killer Robots, *available at* https://www.stopkillerrobots.org/the-problem/ ("Allowing life or death decisions to be made by machines crosses a fundamental moral line. Autonomous robots would lack human judgment and the ability to understand context. These qualities are necessary to make complex ethical choices on a dynamic battlefield, to distinguish adequately between soldiers and civilians, and to evaluate the proportionality of an attack. As a result, fully autonomous weapons would not meet the requirements of the laws of war."); *see also* Heyns, *supra* note 7, at 17 (proposing a moratorium and noting that "[m]achines lack morality and mortality, and should as a result not have life and death powers over humans.").

[69] Vik Kanwar, *Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons*, 2 HARVARD NATIONAL SECURITY JOURNAl 616, 618 (2011) (asking if the role of human combatants recedes, will the respect for humanity be less in either sense? In a "post-human" context of war, where robots take over combat functions, will the connection between these two notions of humanity persist in our overall conception of IHL?).

called for a "preemptive prohibition"[70] of the development of autonomous weapons. HRW argues that "such revolutionary weapons would not be consistent with international humanitarian law and would increase the risk of death or injury to civilians during armed conflict."[71] According to HRW, autonomous weapons "would be unable to follow the rules of distinction, proportionality, and military necessity and might contravene the Martens Clause."[72]

Others have advocated that lethal autonomous weapons should be subject to "meaningful human control,"[73] a term used to "highlight that if the locus of human decision-making and of moral and legal responsibility becomes too far removed from the locus where harm is experienced; if the connection between the two becomes too remote or diffuse or distributed, human control ceases to be meaningful."[74] If a weapon lacking meaningful human control malfunctions and engages human beings, critics argue that "it is possible that no human would be held accountable for those engagements."[75] In a 2016 report focused on the right to assembly and duties of law enforcement, UN Special Rapporteurs Maina Kiai and Christof Heyns recommended

---

[70] Losing Humanity, *supra* note 64, at 1.

[71] *Id*.

[72] *Id.*, at 30.

[73] *Killer Robots: UK Government Policy on Fully Autonomous Weapons*, ARTICLE 36 (Apr. 2013) at 1, *available at* http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf

[74] Maya Brehm, *Meaningful Human Control*, Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 14, 2015) at 4; Michael C. Horowitz and Paul Scharre, *Meaningful Human Control in Weapon Systems: A Primer*, Working Paper, CENTER FOR NEW AMERICAN SECURITY, 11 (Mar. 2015), *available at* http://www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_ Working_Paper_ 031315.pdf.

[75] Horowitz and Scharre, *supra* note 74, at 4.

that "[a]utonomous weapons systems that require no meaningful human control should be prohibited."[76] Some have also argued for "an implicit requirement for human judgment" in IHL[77] and for a "duty" not to delegate the authority to kill to machines.[78]

These concerns are to some extent an outgrowth of the continued "humanization of international humanitarian law."[79] By shifting the focus of regulating war from state security (i.e., "border etc.") to "human security" (i.e., protection of persons and peoples), any weapon that could undermine human security deserves close examination and scrutiny. While this is a legitimate concern, it mischaracterizes the relevance of autonomous weapons in international legal discourse. Indeed, their use brings human security to the center stage because artificially intelligent systems will need to satisfy high levels of confidence before they are fielded. It will require policy makers, military personnel, and developers to flesh out definitions (e.g., proportionality) and develop technologies capable of

---

[76] Maina Kiai and Christof Heyns, *Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies*, United Nations General Assembly, Human Rights Council, A/HR/31/66, at 15 (Feb. 4, 2016).

[77] Peter Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making*, INTERNATIONAL REVIEW OF THE RED CROSS 94, no. 886, 687 (2012), *available at* https://www.icrc.org/eng/assets/files/review/2012/irrc-886-asaro.pdf.

[78] Asaro, *supra* note 78, at 687 ("In particular, there is a duty upon individuals and states in peacetime, as well as combatants, military organizations, and states in armed conflict situations, not to delegate to a machine or automated process the authority or capability to initiate the use of lethal force independently of human determinations of its moral and legal legitimacy in each and every case. I argue that it would be beneficial to establish this duty as an international norm, and express this with a treaty, before the emergence of a broad range of automated and autonomous weapons systems begin to appear that are likely to pose grave threats to the basic rights of individuals.").

[79] Gabriella Blum, *The Individualization of War: From War to Policing in the Regulation of Armed Conflicts*, in AUSTIN SARAT, LAWRENCE DOUGLAS, AND MARTHA MERRILL UMPHREY, LAW AND WAR 48 (Stanford University Press, 2014).

increased precision and discrimination, thus benefiting the objective of maximizing human security.[80]

But like its creator, technology can malfunction, raising questions about who is responsible when things go wrong. In a separate effort focused on the accountability of autonomous weapons, HRW argued that it is "likely that humans associated with the use or production of [autonomous] weapons—notably operators and commanders, programmers and manufacturers—*would escape liability* for the suffering caused by fully autonomous weapons."[81] Others have echoed these concerns.[82] Christof Heyns, the U.N. Special Rapporteur on extrajudicial, summary, or arbitrary executions, questioned whether "lethal autonomous robotics" could be programmed to comply with international humanitarian law and human rights law, and concluded that "their deployment may be unacceptable because no adequate system of legal accountability can be devised."[83] Heyns warned of a "potential accountability gap or vacuum" from the use of such weapon systems.[84] Similarly,

[80] *Cadre Juridique D'un Eventuel Developpement Et Usage Operationnel D'un Futur Systeme D'armes Letal Autonome (Sala)*, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 2016) [hereinafter France Position Paper CCW] *available at* http://www.unog.ch/80256EDD006B8954/(httpAssets)/52B09206029E8FD6C1257F8F0040349E/$file/2016_LAWSMX_CountryPaper_France+LegalFramework.pdf ("En effet, l'utilisation de systèmes d'armes autonomes pourrait réduire les risques pour les civils en prenant des décisions de ciblage plus précises grâce à un calcul plus rapide des informations à leur disposition et des décisions de tir plus contrôlées du fait de l'absence de sentiments négatifs tels la peur, la panique et le désir de vengeance.").

[81] Mind the Gap, *supra* note 66, at 1 (emphasis added).

[82] Jack Beard, *Autonomous Weapons and Human Responsibilities*, 45 GEORGETOWN JOURNAL OF INTERNATIONAL LAW 617, 617 (2014) (noting the "enormous difficulty in assigning responsibilities to humans and states for the actions of these [autonomous] machines grows with their increasing autonomy.")

[83] Heyns, *supra* note 7, at 1.

[84] *Id.*, at ¶ 77.

Yale ethicist Wendell Wallach argues that "attribution and responsibility for the actions of autonomous machines is difficult if not impossible to make."[85]

Others have argued that the authority to legitimately inflict violence in war is exclusive to moral—human—agents.[86] In this context, Maya Brehm argues that "organized violence . . . involves moral agents making moral judgments," which are the result of a "deliberative process of human interaction" and not of "algorithmic calculations."[87] Heather Roff argues that by "hand[ing] over" lethal decision-making to machines, "we jeopardize a moral bedrock of just war theory" because there may be nobody responsible for those decisions.[88] To this end, Wallach and Allen argue that robot systems today only have operational morality, which refers to situations that have been fully pre-coded by the programmer and designer of the system. The more challenging aspect is to develop robots that have functional morality, meaning systems capable of responding to scenarios that were not anticipated (and thus not pre-coded in the traditional sense) by the programmer and designer. In this latter scenario, the robot will need to make ethical decisions alone.[89] As this thesis will

---

[85] WENDELL WALLACH, A DANGEROUS MASTER 219 (Basic Books, 2015) ("Delegating life-and-death decisions to machine is immoral because machines cannot be held responsible for their actions.").

[86] *See e.g.* Heather M. Roff, *Killing in War: Responsibility, liability, and lethal autonomous robots*, in FRITZ ALLHOFF, NICHOLAS G. EVANS, ADAM HENSCHKE (eds.), ROUTLEDGE HANDBOOK OF ETHICS AND WAR 354 (Routledge, 2014) ("the ethical regulation of warfare is premised on the fact that the agents doing the fighting are *moral agents*, i.e., agents to whom responsibility for actions can be attributed.").

[87] Brehm, *supra* note 74.

[88] Roff, *supra* note 86, at 251 ("Furthermore, I contend that when we hand over the decision to target and to fire to a machine, we jeopardize a moral bedrock of just war theory for we move from the central question of 'who is responsible' to 'is there any potential of responsibility?'").

[89] WENDELL WALLACH AND COLLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 9 (Oxford University Press, 2010).

argue in Chapter III, machine learning provides a remedy to the pre-coding paradigm to enable intelligent systems to make ethical decisions without real-time human assistance. Indeed, pessimistic prognoses about autonomous weapon systems in international armed conflicts are based on a misunderstanding of the potential of technology, including machine learning, as well as a failure to fully appreciate the elasticity of international law.[90] To draw an analogy with air combat, "[i]t should be recalled that aeroplanes and drones were first used in armed conflict for surveillance purposes only, and offensive use was ruled out because of the anticipated adverse consequences."[91] Yet the use of aircraft is now typical in military operations and a well-regulated activity under international law.

The fear that autonomous weapon systems will not perform adequately, that they will "act" with impunity, and that they are devoid of morality, may be grounded on human fear of what is unknown or misunderstood, or what is conjured at the intersection of science fictions and science facts. As this thesis argues, the potential for these technologies to adhere to IHL, and for international law to attribute responsibility for their use, is being grossly underestimated by critics.

---

[90] Rodney Brooks, *Artificial intelligence is a tool, not a threat*, RETHINK ROBOTICS BLOG (Nov. 10, 2014) *available at* http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/ ("This all comes from some fundamental misunderstandings of the nature of the undeniable progress that is being made in AI, and from a misunderstanding of how far we really are from having volitional or intentional artificially intelligent beings, whether they be deeply benevolent or malevolent.").

[91] Heyns, *supra* note 7, at 6.

The Technological Distribution

of Violence Is Nothing New

The cry for a ban also ignores the centuries-long distribution of violence tasks between humans and non-human actants. Indeed, the use of automation and autonomy in warfare is "not a new concept."[92] The idea of removing the human body from the immediate infliction of violence has old roots. Take, for example, the bow and arrow. Invented over 10,000 years ago,[93] the bow and arrow automated the propulsion and impact of violence over space and time. Similarly, horses, as autonomous beings,[94] have a long history of being used as "weapon[s]",[95] both literally and figuratively.[96] Virtually all weapons have sought to automate some violent process—or a process for violent ends—while removing the human hand further away from the immediate infliction of violence. Indeed, weapons illustrate the human desire to create unmanned vehicles of violence that position humans as "drivers" without necessarily being "in" the vehicle.[97] For example, the Kettering Bug, developed after the First World War, was a pilotless biplane capable of carrying

[92] CATON, *supra* note 61, at 5.

[93] *Bow and Arrow*, ENCYCLOPEDIA BRITANNICA, *available at* https://www.britannica.com/technology/bow-and-arrow ("The origins of the bow and arrow are prehistoric; bone arrow points dating to 61,000 years ago have been found at Sibudu Cave in South Africa.").

[94] Steve Wise, *'Practical autonomy' entitles some animals to rights*, NATURE (Apr. 25, 2002) *available at* http://www.nature.com/nature/journal/v416/n6883/full/416785a.html (arguing that animals have 'practical autonomy').

[95] MANUEL DELANDA, WAR IN THE AGE OF INTELLIGENT MACHINES 39 (Zone Books, 1991).

[96] CHRIS GRAVETT, TUDOR KNIGHT 29-30 (Osprey Publishing, 2006) (noting that horses used in close combat may have been taught to bite the opponent).

[97] This trend also likely has genetic roots since the removal of the human body away from the doing of violence concurrently removes the body away from exposure to counter-violence.

explosives.[98] Similarly, in the Second World War, the Germans developed a cable-operated vehicle dubbed Goliath capable of carrying explosive ordinances.[99] As early as 1963, DARPA gave the Massachusetts Institute of Technology $2 million to explore "machine-aided cognition," a move that jump-started research in AI.[100] In 1984, in a document called "Strategic Computing", the Pentagon revealed its intention to create "autonomous weapons systems capable of fighting wars entirely on their own."[101]

In *War in the Age of Intelligent Machines*, Manuel DeLanda traces the evolution of the "war machine" and argues that "[f]or centuries, military commanders have dreamed of eliminating the human element from the battlefield."[102] To illustrate his argument, DeLanda identified three phases in projectile-based production of violence: (i) the propulsion stage, covering everything up to the firing of a weapon (ii) the ballistic stage, covering its navigation, and (iii) the impact stage.

The propulsion stage encompasses every task necessary to allow the projectile to be propelled out of the weapon. In particular, three components make up the propulsion stage: fueling (i.e., loading a weapon); ignition (i.e., the triggering act);

---

[98] Wagner, *supra* note 39, at 5-6.

[99] *Id*.

[100] Ty McCormick, *Lethal Autonomy: A Short History*, FOREIGN POLICY (Jan. 24, 2014) *available at* http://foreignpolicy.com/2014/01/24/lethal-autonomy-a-short-history/.

[101] DELANDA, *supra* note 95, at 128.

[102] DELANDA, *supra* note 95, at 128; Geoffrey S. Corn, *Autonomous weapons systems: managing the inevitability of 'taking the man out of the loop'*, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, 209 ("Military leaders will constantly seek both the means (weapons) and the methods (tactics) of warfare to maximize their full-spectrum dominance over their adversaries.").

and guidance (i.e., the imparting of the projectile).[103] The 14th century hand canon, one of the first firearms, was unable to coordinate these three functions and depended heavily on human agents to coordinate them. The smoothbore tube served as the "only guidance mechanism, so that the rest of the process depended on human marksmanship"[104] and the human gunner had to light the fuse to fire a weapon. The fueling component was enhanced by the development of the metallic cartridge and breech-loading, which in contrast to its predecessor, known as "muzzle loading," revolutionized the operation of the firearm.[105] In 1424, the ignition process was simplified with a mechanical device—the matchlock—removing the human gunner's involvement in that process.[106] The manufacture of weapons also experienced a wave of automation with the replacement of individual gunsmiths with military engineers[107] "who began the military rationalization of labor in American armories and arsenals."[108] The standardization of labor in the production of military equipment effectively automated a process previously held by heavily specialized—manually intensive—labor.[109] The Industrial Revolution precipitated this process.

---

[103] DELANDA, *supra* note 95, at 25.

[104] *Id.*

[105] *Id.*, at 27.

[106] *Id.*, at 25.

[107] *Id.*, 29.

[108] *Id.*, at 56.

[109] *Id.*, at 35.

In the ballistic stage, DeLanda sought to examine the evolution of "events influencing the trajectory of the missile in flight."[110] Here, the role of information technology becomes increasingly relevant. DeLanda argues that early computers, in the form of mechanical calculators and people using those calculators, were deployed to create artillery range tables to assist gunners in the calculation of accurate missile trajectories.[111] With the advent of the Differential Analyzer (a mechanical analogue computer), the people making these calculations were taken "out of the loop" as the process was automated. As DeLanda explains:

> The next stage in this process would involve transferring the gunner's calculating skills to the launching platform, to take him out of the decision-making loop. The artillery range tables produced by automatic devices "were programmed into analog computers called gun directors which took over the job of calculating trajectories from the human antiaircraft gunner. Eventually the gun directors were connected to radar systems, channeling information about target location directly to control the guns . . . . The gunner had to predict how far ahead of a fast-moving plane he had to aim so that the trajectories of his missile and the plane would intersect at the right point. This job of prediction was taken out by servomechanism (feedback-based) devices."[112]

The automation of the ballistic stage was further entrenched with the development of fire and forget weapons, aided by heat-seeking and computer controlled technology.[113] As DeLanda argues, "[a]rtificial [i]ntelligence would create the techniques necessary for building autonomous weapons systems endowed with

---

[110] *Id.*, at 39.

[111] *Id.*, at 35.

[112] *Id.*, at 46.

[113] *Id.*, at 35.

predatory capabilities of their own."[114]

The third stage, which DeLanda calls the "impact stage," overlaps with the prior stages and focuses on automating both the offensive and defensive aspects of the impact of projectile-based violence. Here, DeLanda uses the radar to illustrate how walls and fortresses were rendered obsolete by missile technology, which in turn paved the way for a more expansive notion of a wall, this time in an electromagnetic form—the radar curtain—[115] and developments in stealth technology shortly ensued. More recently, unmanned drones further removed human beings from the site of impact.

The "iCombat world,"[116] defined by a deeper collaboration between humans and technology, is thus the product of a centuries-long redistribution of violence tasks within networks of human violence. As DeLanda put it:

> The robotic predator…may be seen as the culmination of the long "bridging" process started by electrical engineers and ballisticians in World War I, to channel scientific know-how into the creation of missiles and guns ever-less dependent on human skill for their performance…The efforts of military institutions to get humans out of the loop have been a major influence in the development of computer technology. The birth of autonomous weapons systems, of war games played by automata, of production systems that pace and discipline the worker, all are manifestations of this military drive...[117]

The removal of human involvement from the *propulsion*, *ballistic,* and *impact* loops of violence has thus been an ongoing objective in networks of human

[114] *Id.*at 46.

[115] *Id.*, at 51 ("The next stage in the development of the wall occurred when offense technology created a new delivery vehicle, the bomber plane, forcing the fortress to dematerialize into the electronic radar curtain.").

[116] Pellerin, *supra* note 45.

[117] DELANDA, *supra* note 95, at 46, 177.

violence. The end result is not a desire to entrust non-human actants with sole decision-making powers over life and death, but to *incorporate* technology capable of automation and autonomy within human networks to deliver force on the enemy.

Of course, autonomy in robotic machines raises additional issues that are distinct from the automation of producing weapons[118] and automating independent components of how they work. Entrusting a weapons system relative independence over targeting and firing decisions raises vivid concerns about predictability, competence, and accountability. Yet these weapons are merely a method of human violence by other means.

Revised Taxonomy:
Focus on Intelligence, Not Autonomy

Many of the concerns about autonomous weapons systems stem from a perceived cessation of human control. This perception is in part caused by terminological uncertainties. Systems capable of autonomous decision-making in armed conflicts have been described in multiple forms, including lethal autonomous robots ("LARs"),[119] lethal autonomous weapon systems ("LAWS"),[120] Killer Robots,[121] and autonomous military robots.[122] These taxonomies create overtures for

[118] *See id.*, at 177 ("In the early nineteenth century, the American military began to transform the mode of operation of its armories in order to produce firearms with perfectly interchangeable parts. To achieve this goal, they introduced methods for the routinization and standardization of labor.")

[119] *See* Roff, *supra* note 86.

[120] *See* Heyns, *supra* note 7.

[121] *See* Losing Humanity, *supra* note 64.

mischaracterizations and hyperbolic constructions that do not reflect the true nature or the potential of artificial intelligence. Indeed, the use of the term "autonomy" complicates our understanding of the technology and undermines the reality of human design and human use. There are a number of reasons why the term "autonomy" should be replaced with "intelligence." Here, this thesis explains why the term Violent Intelligent Systems ("VIS") is more fitting to describe violent emerging technologies capable of autonomous operation.

Autonomy has been described in a variety of ways, and no uniform or "stable"[123] definition exists in the technological, philosophical,[124] or legal spheres.[125] Indeed, even in a discipline that conjures notions of precision (i.e., engineering), "autonomy means several different things."[126] In the military context, autonomous weapon systems have been described by the DoD as a "weapon system that, once activated, can select and engage targets without further intervention by a human operator."[127]

---

[122] Merel Noorman and Deborah G. Johnson, *Negotiating autonomy and responsibility in military robots*, 16 ETHICS AND INFORMATION TECHNOLOGY 51, 51 (2014).

[123] Nehal Bhuta and Starvos-Evdokimos Pantazopoulos, *Autonomy and uncertainty: increasingly autonomous weapons systems and the international legal regulation of risk*, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, at 285 (Cambridge University Press, 2016) ("no stable consensus exists concerning the meaning of autonomy or of autonomy in weapons systems").

[124] *Autonomy in Moral and Political Philosophy*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY, *available at* http://plato.stanford.edu/entries/autonomy-moral/ ("the concept of autonomy is the focus of much controversy and debate").

[125] Lewis, Blum, and Modirzadeh, *supra* note 6, at iii ("Largely, the discourse to date has revolved around a concept that so far lacks a definitional consensus… On one end of the spectrum, an AWS is an automated component of an existing weapon. On the other, it is a platform that is itself capable of sensing, learning, and launching resulting attacks.").

[126] DAVID A. MINDELL, OUR ROBOTS, OURSELVES 11 (Viking, 2015).

[127] Directive Number 3000.09, *supra* note 35, at 13-14.

It may be easier to describe autonomy by reference to what it is not. Autonomy is different from remote controlled systems and automatic systems.[128] Whereas automatic systems are "fully preprogrammed and act repeatedly and independently of external influence or control" and can be 'self-steering or self-regulating' but cannot 'define' or 'dictate' their own paths,"[129] autonomous systems on the other hand are "self-directed toward a goal in that they do not require outside control,"[130] *and* are "able to make a decision on a set of rules and/or limitations"[131] based on information that they deem important to the decision process.[132] However, making such decisions about rules is accomplished by way of "programming platforms and approaches and a series of information inputs over repeated experiences of interactions." [133] Thus, while a landmine is automatically indiscriminate, an autonomous weapon, powered by artificial intelligence, would be capable of discrimination based on its programming parameters.

---

[128] Wagner, *supra* note 39, at 9 ("The different types of unmanned systems can be usefully grouped into three different categories, although these classifications are more realistically described as existing on a spectrum that moves from human-controlled systems towards full autonomy: remotely operated systems, automated systems, and systems that operate autonomously.")

[129] Roff, *supra* note 86, at 353.

[130] *Id.*, at 353.

[131] *Id.*, at 353.

[132] *See also Unmanned Systems Integrated* Roadmap *FY2011-2036,* United States Department of Defense, Ref. No. 11-S-3613, 2011, *available at* http://www.acq.osd.mil/sts/docs/Unmanned%20 Systems%20Integrated%20Roadmap%20FY2011-2036.pdf ("In its 2011 Roadmap, the DoD argues that automatic systems are fully preprogrammed and act repeatedly and independently of external influence or control. They are able to follow a predefined path while compensating for small deviations caused by external disturbances. In contrast, autonomous systems are self-directed toward a goal in that they do not require outside control, but rather are governed by laws and strategies that direct their behavior. Their behavior in response to certain events is not fully specified or pre- programmed. According to the DoD '[a]n autonomous system is able to make a decision based on a set of rules and/or limitations. It is able to determine what information is important in making a decision'").

[133] Roff, *supra* note 86, at 353.

The notion that autonomy implies no human control is arguably founded on the misconception that autonomy implies *full* autonomy. But full autonomy is a techno-mythic fantasy. Autonomy is best understood to operate along a "continuum"[134] or a "spectrum"[135] of capabilities.[136] For example, the DoD's definition of autonomous weapon systems includes *human-supervised* autonomous weapons systems "that are designed to allow human operators to override operation" of the weapon.[137] Autonomy is a capability across different degrees and levels that

---

[134] Tim McFarland and Tim McCormack, *Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes?*, 90 INT'L L. STUD. 361, 369 (2014) ("Machine autonomy is a capability that exists within a continuum rather than at discrete levels, ranging from complete human control over some operations to complete computer control. Military and civilian research organizations have proposed many different taxonomies of autonomous capability.").

[135] Wagner, *supra* note 39, at 9 ("The different types of unmanned systems can be usefully grouped into three different categories, although these classifications are more realistically described as existing on a spectrum that moves from human-controlled systems towards full autonomy: remotely operated systems, automated systems, and systems that operate autonomously."); *Canadian Food for Thought Paper: Mapping Autonomy*, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW), 2006, *available at* http://www.unog.ch/80256EDD006B8954/(httpAssets)/C3EFCE5F7BA8613BC1257F8500439B9F/$file/2016_LAWS+MX_CountryPaper+Canada+FFTP1.pdf ("Levels of autonomy vary according to a variety of factors. It is likely more useful to think in terms of a spectrum of autonomy, with the level of autonomy closely tied to a system's technology and capabilities, operational environment, and chosen task, rather than merely the qualities of the system itself.").

[136] HRW has developed a tripartite taxonomy that seeks to situate the human along the following spectrum: (i) "Human-in-the-Loop Weapons: Robots that can select targets and deliver force only with a human command," (ii) "Human-on-the-Loop Weapons: Robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions," and (iii) "Human-out-of-the-Loop Weapons: Robots that are capable of selecting targets and delivering force without any human input or interaction." Losing Humanity, *supra* note 64, at 2. While easy to follow, a weakness of HRW's taxonomy is that it employs hidden assumptions about where the human is (and be) situated. It also makes presumptions about what a "human" is and what terms like "human input" and "human command" encompass. For example, couldn't "human command" include commands by humans expressed through code? Similarly, the "Human out of the Loop" definition is based on the assumption that there was no "human input" in selecting a target, but wholly ignores that human-made code is responsible for modelling the selection of targets and the decision-making process of delivery force, if any. *See also* Noel Sharkey, *Staying in the loop: human supervisory control of weapons*, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, at 27 ("human engages with and selects target and initiates any attack; program suggests alternative targets and human chooses which to attack; program selects target and human must approve before attack; program selects target and human has restricted time to veto; program selects target and initiates attack without human involvement").

[137] Directive Number 3000.09, *supra* note 35, at 13-14.

adjust depending on context.[138] It can refer to independent system capabilities, including perception, planning, learning, interacting, language understanding, and multi-agent coordination.[139] But it also goes beyond computational autonomy. Indeed, "a robot cannot be said to be truly autonomous unless it has energy autonomy . . . there seems little point in building a smart robot that 'dies' when its battery runs out."[140] We are far from building a system that does not require humans for energy, healing, and reproduction.[141]

The overall consensus on how to best describe autonomy is that it is "not as a discrete property of an object or system, but rather as a relationship between a system and its operator that may vary across the spectrum of different degrees of system autonomy."[142] As Johnson and Norman note:

> [M]aking robots autonomous in various ways means that human actors have different kinds of control. Human actors exert their influence as

---

[138] *The Role of Autonomy in DoD Systems*, Task Force Report, UNITED STATES DEPARTMENT OF DEFENSE, DEFENSE SCIENCE BOARD (Jul. 2012) *available at* http://fas.org/irp/agency/dod/dsb/autonomy.pdf at 23 ("The competing definitions for autonomy have led to confusion among developers and acquisition officers, as well as among operators and commanders. The attempt to define autonomy has resulted in a waste of both time and money spent debating and reconciling different terms and may be contributing to fears of unbounded autonomy. The definitions have been unsatisfactory because they typically try to express autonomy as a widget or discrete component, rather than a capability of the larger system enabled by the integration of human and machine abilities.").

[139] Roff, *supra* note 86, at 353 ("In a 2012 task force report, the DSB identified "six key areas in which advances in autonomy would have significant benefit to the unmanned system: [perception, planning, learning, human-robot interaction, natural language understanding, multiagent coordination.]").

[140] ALAN WINFIELD, ROBOTICS: A VERY SHORT INTRODUCTION 49 (Oxford University Press, 2012).

[141] *Id.*, at 8. This is not to mean that investments are not already being made in these specific capabilities. Indeed, in the context of energy management, the defense contractor Raytheon has developed small robots named Hercules and Athena that run on solar power but are programmed to stay out of the light, a conflict that pushed the robots to develop "hunger" for energy and better life-death management skills. Davenport, *supra* note 46. Harvard researchers have also built an 'octobot', a "self-contained" robot that is powered by chemical reactions and does not need batteries. Jonathan Webb, *Pneumatic octopus is first soft, solo robot*, BBC NEWS (Aug. 25, 2016) *available at* http://www.bbc.com/news/science-environment-37169109.

[142] CATON, *supra* note 61, at 2.

> they choose the mathematical and probabilistic models that will guide the behavior of the robotic system; as they formulate restrictions on the conditions for use and specify and verify the levels of reliability and predictability that robotic systems need to exhibit. Designers, developers, human operators as well as managers, regulators and policy makers, thus, set constraints on what robotic systems can and cannot do.[143]

The emphasis on autonomy also misunderstands military operational realities. The absence of human liaison with a weapon system "contradicts the need for permanent and accurate situation awareness and the operational control by the commander."[144] Commanders generally have no interest in deploying weapon systems that "they cannot control."[145] Even malevolent regimes and non-state groups that desire to use these weapon systems for nefarious purposes will need to program them to meet their goals.

Far from a "Frankensteinian fantasy,"[146] Violent Intelligent Systems are thus designed, programmed, and deployed by humans to achieve human objectives. There is no "[d]ehumanization" of lethal decision-making.[147] While "[t]he crux of full autonomy is a capability to identify, target, and attack a person or object without

---

[143] Noorman and Johnson, *supra* note 122, at 59-60.

[144] France Position Paper CCW, *supra* note 80.

[145] Horowitz and Scharre, *supra* note 74, at 8 ("A military's tolerance for risk could vary considerably across cultures and strategic positions . . . . A desire for a battlefield advantage could push militaries to build weapons with high degrees of autonomy that diminish human control, particularly if they see such weapons as necessary to confront emerging threats or to keep pace with other militaries").

[146] Brad Allenby, *What Human Rights Watch's "Case Against Killer Robots" Gets Wrong About Military Realities,* SLATE (Nov. 20, 2012), *available at* http://www.slate.com/blogs/futuretense/2012/11/20/human rights watch_s_case_against killer robotsreport misunderstands _the.html.

[147] Asaro, *supra* note 78, at 687-709; Heyns, *supra* note 7, at 17 ("Delegating this process dehumanizes armed conflict even further").

human interface,"[148] even "a fully autonomous system is never entirely human-free"[149] as "human beings will inevitably be involved, either in overseeing the operation of the weapon, or at least in producing and programming the weapon systems."[150] Indeed, it is highly unlikely for there to be fully autonomous weapon systems for the systems will always be subjected to the socialization of design and the parameters of code.

The suggestion that autonomy implies humans are "out of the loop" should therefore be revised. As this thesis argues, Violent Intelligent Systems, and the code that powers them, are expressions of human violence that cannot be disconnected from its human origin. Autonomous systems execute tasks "on the designer's behalf."[151] Autonomy is thus best conceived as "a *human-designed* means for transforming data sensed from the environment into purposeful plans and actions."[152] These weapon systems "are not conducting automated warfare—people are still

---

[148] Michael N. Schmitt, *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, HARV. NAT'L. SEC. J. FEATURES (2013), *available at* http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf at 4

[149] Schmitt and Thurnher, *supra* note 54, at 235 ("Either the system designer or an operator would at least have to program it to function pursuant to specified parameters, and an operator would have to decide to employ it in a particular battlespace.").

[150] Marco Sassóli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'L. L. STUD. 308, 309 (2014).

[151] Eduardo Alonso, *Actions and agent*, in KEITH FRANKISH AND WILLIAM M RAMSEY (eds.), THE CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 235 (Cambridge University Press, 2014) ("By autonomy researchers mean the ability of the systems to make their own decisions and execute tasks on the designer's behalf.").

[152] MINDELL, *supra* note 126, at 12.

inventing, programming, and operating [them]."[153] In other words, humans will "always . . . define how this autonomy will function."[154]

A focus on autonomy has also led for calls to create a legal requirement for "meaningful human control" over their use. Two schools of thought emerge in this context. The first does not propose an additional requirement but "rather a principle for the design and use of weapon systems in order to ensure that their use can comply with the laws of war."[155] As the Center for a New American Security points out, this "way of thinking about meaningful human control starts from the assumption that the rules that determine whether the use of a weapon is legal are the same whether a human delivers a lethal blow directly, a human launches a weapon from an unmanned system, or a human deploys an autonomous weapon system that selects and engages targets on its own." As part 3 of this thesis argues, international law already provides for this safeguard. However, a more "maximalist" way of thinking about "meaningful human control" requires more than what current international law provides. Here, "meaningful human control" is a separate and additional requirement that must be satisfied under the laws of war, though its contents are unclear.

Further, "meaningful" is an open textured concept that is difficult, if not impossible, to define,[156] and the term "human control" invites the question of control

[153] *Id.*, at 13.

[154] Sassóli, *supra* note 150, at 323-324 ("I do not think that the possession of autonomous decision making capacity breaks the causal chain allowing attribution and responsibility, because I assume that it is always humans who define how this autonomy will function.").

[155] Horowitz and Scharre, *supra* note 74, at 7.

[156] CNAS surveyed the positions made by those advocating for "meaningful human control" and concluded that there are three "essential components" of the proposed requirement. First, human operators must be able to make "informed, conscious decisions about the use of weapons." Second,

"at what level" [157] and in what form? Efforts should be redirected to focus instead on whether an intelligent system is capable of meeting the requirements of IHL and other applicable rules (i.e., rules of engagement). Whether the human is holding the trigger or marking the target is hardly relevant if the system, powered by artificial intelligence, cloud networking, and advanced robotics, is capable of adhering to legal parameters. As one commentator put it, "[e]ven seemingly indisputable calls for a first principle of 'meaningful human control' mistake the issue, which is lessening the harms of armed conflict within the law by the means that are the most effective."[158] Indeed, as this thesis argues, a VIS may act more intelligently than human soldiers and thus be able to better adhere to IHL.[159]

Ultimately, decisions to develop and deploy autonomous weapon systems are firmly "concentrated in human hands."[160] Human judgment need not exclusively mean contemporaneous human supervision or control, but it also "includes the ability to decide when to let a machine operate autonomously—in other words, to forfeit [human] judgment, at least temporarily—and when to maintain [human] control over

---

human operators have "sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action." Third, "[t]he weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon." Horowitz and Scharre, *supra* note 74, at 4.

[157] Horowitz and Scharre, *supra* note 74, at 15.

[158] Kenneth Anderson, Daniel Reisner, and Matthew Waxman, *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 INT'L L. STUD. 386, 401 (2014).

[159] Until artificial intelligence and robots are capable of reproducing, it is clear that human control will pervade the design, programming, and operation of VIS with autonomous capabilities. The code that runs through the semi-conductors and software of the systems are products of human will and expressions of human judgment – and thus control – in electronic form.

[160] PAUL DUMOUCHEL AND LUISA DAMIANO, VIVRE AVEC LES ROBOTS 201 (Seuil, 2016) ("la devolution de la capacite de choisir a des systems autonomes intensifie la concentration entre les mains de quelques agents humains seulement de la capacite politique et morale de decider").

the machine."[161] In sum, the "decision to kill . . . is always taken by human beings [who] set out the ethical constraints to autonomous systems' behavior . . . . Military robots are actually mechanical slaves that can only obey [human commands]."[162]

Accordingly, the focus on autonomy should be replaced with a focus on "intelligence." The latter better captures both the technological realities and policy sensitivities surrounding the use of weapons capable of autonomous operation. Technologically, unless systems are able to autonomously reproduce and become self-sufficient, it is misleading to speak exclusively of "autonomy" in weapon systems. A shift from autonomy to intelligence is better suited to describe systems that are capable of running thousands if not millions of calculations in complex and uncertain situations to achieve human objectives.[163] The code that powers the weapons is a product of human design. Even algorithms capable of machine learning are tied to the human-made code that dictates their parameters. So when we look under the hood, code has no real autonomy. The physical hardware that materializes code's instructions (the gun subsystems, the sensors, etc.) are all designed by human beings and human institutions and do not "magically" animate themselves.[164]

---

[161] Dan Saxon, *In the Context of the Design and Use of Autonomous Weapon Systems, What is Judgment?*, Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 2016).

[162] DUMOUCHEL AND DAMIANO, *supra* note 160, at 203.

[163] Alonso, *supra* note 151, at 236 ("[I]ntelligence and learning are tightly tied in domains where autonomous agents must make decisions with partial or uncertain information: that is, in domains where agents learn without supervision and without the luxury of having a complete model of the world.").

[164] Noorman and Johnson, *supra* note 122, at 59 ("Human actors exert their influence in at least three ways. First, much like in the Phalanx case, developers and designers delimit the problem that the robotic system is intended to solve and thus set constraints on its behavior . . . . A second way that human actors exert their influence on autonomous robots that are somehow more than automatic systems, is through norms and rules; even in future systems, norms and rules will still govern the

Intelligence better encompasses the capacity of code, hardware, and human designers and operators *to interact* in a network to organize and produce violence. It also reflects the reality that artificial intelligence is the driving force of these systems.

From a policy standpoint, the focus should not be whether "a system is *too* autonomous" but rather "is the system sufficiently intelligent to execute the task at hand in compliance with the law?" Autonomy should not be the yardstick by which we measure weapon systems capable of autonomous operation. Rather, the question should be whether intelligent systems can operate *responsibly* in accordance with legal requirements. Autonomous beings (humans-included) can act irresponsibly, can fail to learn from their experiences, and can fail to accomplish stated goals. Intelligence, on the other hand, requires learning from and adapting to the operating environment. Indeed, "an [AI] agent can hardly be called intelligent if it is not able to perform well when situated in an environment different from (yet in some ways similar to) the one it was originally designed for."[165] If we manage to build a completely autonomous system that fails to act intelligently and responsibly, would not we want to program it with some intelligence instead? Should not these weapon systems be judged by their intelligence (or lack thereof), rather than their autonomy (or lack thereof)?

---

behavior of autonomous systems . . . . A third way that human actors exert their influence on autonomous robots has to do with predictability. Conceiving of autonomous robotic systems as somehow more flexible and nondeterministic than conventional automation calls for an increased emphasis on reliability and trust in technology, and the need to develop better methods for verification and validation (V&V).").

[165] Alonso, *supra* note 151, at 236 ("[I]ntelligence and learning are tightly tied in domains where autonomous agents must make decisions with partial or uncertain information: that is, in domains where agents learn without supervision and without the luxury of having a complete model of the world.").

Further, as a medium of artificial intelligence, code may allow for a more rigorous translation of IHL principles into military operation and thus better ensure that VIS operate *responsibly*, regardless of the system's autonomous capabilities. In other words, whereas intelligence can be regulated, autonomy implies uncertainty and irregulability.[166] Moreover, terminologically, VIS suggest that these are creations of human design and engineering,[167] rather than the product of some bizarre Frankensteinian alchemy.[168]

The hostility toward autonomous weapons is also exacerbated by the association of autonomy with lethality. Terms like "killer robots" or "lethal autonomous robots" conjure a sense of uncontrolled, *Terminator*-like,[169] machines incapable of abiding by human commands—or worse, threatening the very existence of the human species.[170] Lethality is a capacity—and potentially a consequence—of violence. The term violence better encompasses the types of uses of weapons with autonomous capabilities, which may include lethal and non-lethal objectives. Violence also reflects the technological capability of developing a weapon that neutralizes, and not necessarily kills, a target. Indeed, "the [U.S.] military is already

---

[166] I am referring to non-human autonomous agents here.

[167] MINDELL, *supra* note 126, at 220 ("[A]ny supposedly intelligent system was designed by people").

[168] In this context, the language of technology must be aligned with the language of the law.

[169] *See* BROOKS, *supra* note 4, at 135 ("The term does tend to have a chilling effect even upon those harboring a soft spot for R2-D2 and WALL-E. I'm less concerned, however: not because I'm fond of killer robots, but because I'm inclined to think ethicists and rights advocates are far too generous in their assumptions about human beings.").

[170] Ted Greenwald, *Does Artificial Intelligence Pose a Threat*, WALL STREET JOURNAL (May 10, 2015) *available at* http://www.wsj.com/articles/does-artificial-intelligence-pose-a-threat-1431109025 (reporting on those that believe that AI poses an existential threat to mankind).

experimenting with a number of technologies that can incapacitate or control enemies without causing injury or death."[171] The DoD has been developing "active denial technology" which shoots a "beam of radio frequency millimeter waves toward a specific area," effectively paralyzing the target "with no permanent ill effects" and thus giving the military more time to suppress ambushes and identify targets from civilians.[172] Violence can thus be used for primarily non-lethal purposes, such as deterrence, or to protect vulnerable sites, such as schools, hospitals, and culturally significant sites.[173]

In the context of autonomous weapons, the word "robot" embodies the idea of a separate technological unit, further personifies the technology, and thus intensifies the focus on autonomy.[174] By contrast, the word "system," meaning "a group of

---

[171] BROOKS, *supra* note 4, at 139.

[172] *Id.*, at 139-140; *Active Denial Technology Fact Sheet,* U.S. DEPARTMENT OF DEFENSE, Non-Lethal Weapons Program (May 11, 2016) *available* at http://jnlwp.defense.gov/Press-Room/Fact-Sheets/Article-View-Factsheets/Article/577989/active-denial-technology-fact-sheet/ ("Active Denial Technology produces a focused beam of directed energy to provide our troops a non-lethal option to stop, deter and turn back suspicious individuals with minimal risk of injury. Active Denial Technology is designed to protect the innocent, minimize fatalities and limit collateral damage across the range of military operations.").

[173] Lucas Bento, *Could Science Defeat Terrorism? Using Robots to Hunt Down ISIS*, THE DIPLOMAT (Aug. 31, 2015) ("Killer robots would be particularly useful against groups like ISIS, where political costs are too high for major military powers to put boots on the ground, and political momentum too low to justify human military intervention to protect sites of cultural importance."); Thiago Velozo and Lucas Bento, *ISIS Is Destroying Priceless Artifacts. Here's How to Stop Them.*, THE DIPLOMAT (Mar. 17, 2015) (noting how technology can help preserve cultural artifacts).

[174] The use of the term "robot", while easy to understand and technically correct in certain contexts, carries significant connotations of personhood or fictional imagery of an animated 'being'. *See e.g. Mind the Gap*, *supra* note 66, at 2 ("The autonomous nature of killer robots would make them legally analogous to human soldiers in some ways, and thus it could trigger the doctrine of indirect responsibility, or command responsibility."). The word "comes from an Old Church Slavonic word, rabota, which means servitude of forced labor." The word robot was introduced and popularized by Czech playwright Karel Čapek, who used the word in his 1920 science-fiction play, *R.U.R.*, or *Rossum's Universal Robots*, which involved a factory that made artificial people to work for humans. *See* Ira Flatow, *Science Diction: The Origin of the Word 'Robot',* NPR (Apr. 22, 2011), *available at* http://www.npr.org/2011/04/22/135634400/science-diction-the-origin-of-the-word-robot.

related parts that move or work together,"[175] is better suited to describe the interactive complexities of the technology. Indeed, the term "weapon system" is already the term of choice among some stakeholders (though they also pair it with "autonomous").[176]

A Violent Intelligent System can operate sub-systems that include manned or unmanned platforms, munitions, or sub-munitions that can function autonomously or semi-autonomously.[177] For example, in one mission the VIS may identify and target autonomously, but only fire after human real-time approval. The same VIS in other missions may carry weapons that do the firing without human approval. To this end, commentators note that "[t]here is no reason to suppose that all operations of an autonomous system will be subject to the same degree of human oversight, and a system may be operating at several different levels of autonomy simultaneously with respect to different tasks."[178]

As artificial intelligence gains prominence in both traditional and cyber warfare, it will increasingly become integrated with traditional conceptions of space. To this end, the US military has been an early adopter of the internet of things ("IoT") and is seeking to expand its applications on the battlefield. This will further entrench

[175] Mariam-Webster Dictionary, *available at* http://www.merriam-webster.com/dictionary/system.

[176] *See e.g.* Directive Number 3000.9 *supra* note 35; Noorman and Johnson, *supra* note 122.

[177] Directive Number 3000.9, *supra* note 35, at 3 ("including manned or unmanned platforms, munitions, or sub-munitions that function as semi-autonomous weapon systems or as subcomponents of semi-autonomous weapon systems"); McFarland and McCormack, *supra* note 134, at 369 ("This consideration also applies to a 'system of systems' scenario wherein nominally separate systems with different levels of autonomy, such as an intelligence, surveillance and reconnaissance (ISR) system and a weapons system, communicate directly.").

[178] McFarland and McCormack, *supra* note 134, at 369.

the collaboration of human and non-human actants, as objects will become sources of information that can shape decision-making by both humans and weapon systems capable of autonomous operation.[179]

In the final analysis, the distribution of violence between humans and non-human actants is a product of human political and sociotechnical design.[180] While the mechanics that power modern weapons are beyond natural human abilities, they are nonetheless man-made creations. Fighter jets, drones, battleships, cruise missiles, and other means of warfare, are not standalone objects, dangling in the void, but deliberately placed components in networks of human violence.[181]

The "human design" of violence is not only to be understood as the actual design of the weapon systems themselves, but more perhaps importantly, the wider strategies that seek to engineer methods of violence that incorporate increased artificially intelligent capabilities while seeking to remove the human body away from the immediate infliction of violence. Human design, in turn, entails human

---

[179] Joe Mariani, Brian Williams, and Brett Loubert, *Continuing the march: the past, present, and future of the IoT in the military*, DELOITTE UNIVERSITY PRESS (Aug. 6, 2015) *available at* http://dupress.com/articles/internet-of-things-iot-in-military-defense-industry/.

[180] Noorman and Johnson, *supra* note 122, at 61 ("The ascription of responsibility is therefore an integral part of the development and design of robots. Delegation of responsibility to human and non-human components is a sociotechnical design choice, not an inevitable outcome of technological development.").

[181] Noorman and Johnson, *supra* note 122, at 59 ("Human actors exert their influence in at least three ways. First, much like in the Phalanx case, developers and designers delimit the problem that the robotic system is intended to solve and thus set constraints on its behavior… A second way that human actors exert their influence on autonomous robots that are somehow more than automatic systems, is through norms and rules; even in future sys- tems, norms and rules will still govern the behavior of autonomous systems . . . . A third way that human actors exert their influence on autonomous robots has to do with predictability. Conceiving of autonomous robotic systems as somehow more flexible and nondeterministic than conventional automation calls for an increased emphasis on reliability and trust in technology, and the need to develop better methods for verification and validation (V&V).").

responsibilities that are encompassed by well-established principles of international law.

Chapter III

Human Responsibility

Moderating War through Legal and Practical Responsibilities

War is one of the most destructive human enterprises.[182] The ability to moderate[183] war's brutality and to attribute responsibility for its consequences informs the twin pillars of legal responsibility: acting responsibly,[184] and being responsible for one's acts. Legal responsibility, then, can be divided as "responsible-ability" and "responsible-liability," with the former governed by *jus ad bellum* and *jus in bello* principles[185] and the latter by principles of international criminal law ("ICL") and state responsibility. These mechanisms are important because "[w]ithout the promise of accountability, deterrence and prevention are reduced, resulting in lower protection of civilians and potential victims of war crimes."[186]

---

[182] MICHAEL P. JASINSKI, SOCIAL TRUST, ANARCHY, AND INTERNATIONAL CONFLICT 3 (Palgrave Macmillan, 2011) (describing war as "the most senseless and destructive of human enterprises"); *see also* CARL VON CLAUSEWITZ, ON WAR 242 (Wordsworth Editions, 2000) ("The destruction of the enemy's military force is the leading principle of war").

[183] JONATHAN CROWE AND KYLIE WESTON-SCHEUBER, PRINCIPLES OF INTERNATIONAL HUMANITARIAN LAW 2 (Edward Elgar, 2013) (arguing that IHL seeks to moderate the effects of warfare and noting that, as such, IHL "aims to ensure respect for the most basic human values, such as dignity, community, and freedom from suffering.").

[184] At bottom, IHL was developed to address a consensus that humans were incapable of conducting themselves "humanely" in armed conflicts. After all, law does not exist to remedy a vacuum.

[185] Principles of international human rights law may also apply.

[186] Heyns, *supra* note 7, at 14; Mind the Gap, *supra* note 66, at 14 ("One of the primary reasons to hold individuals accountable is to deter harmful behavior. If individuals are punished for unlawful acts, they may be less likely to repeat them. Holding offenders responsible can also discourage future infractions by other actors, who fear being punished in the same way. According to Dinah Shelton, author of the treatise Remedies in International Human Rights Law, "[d]eterrence . . . is

In addition to legal responsibility, the notion of responsibility can also refer to practical responsibilities. In functional terms, practical responsibilities are distributed among multiple actants in networks of human violence. For example, the practical responsibility for a cruise missile to detonate upon impact lies with the engineering system embodied within the missile. Legal responsibility for that impact, however, lies with human actors that engineered the attack. Of course, practical and legal responsibilities inform each other. Indeed, if a weapon is deemed practically incapable to discriminate between a civilian and a military target, humans are required under IHL not to use that weapon, and any such use would make human actors and human institutions responsible for it. Thus, while practical responsibilities may be shared between human actors and non-human actants, legal responsibilities are not similarly distributed.[187] As this thesis argues, only human beings and human institutions share legal responsibilities for the participation of non-human actants, such as VIS, in the delivery of violence in international armed conflicts.

The pursuit of mitigating[188] war's brutality has old roots.[189] Sumerian law prescribed "specific rules" to govern wartime conduct, such as the granting of

---

assumed to work because rational actors weigh the anticipated costs of transgressions against the anticipated benefits." She adds that "[d]eterrence literature also shows a correlation between the certainty of consequences and the reduction of offences." For deterrence to have the maximum effect, potential offenders must have advance notice of the prospect of accountability so that they can consider the consequences before they act. Public assurances that steps are being taken to diminish the likelihood of new offenses can also provide consolation to victims and society.")

[187] Liu, *supra* note 65, at 629 ("Responsibility in law is a concept that has several disparate dimensions. Thus, although it may be possible for a machine to be responsible in a strictly causal sense for the production of specific results or outcomes, these are not necessarily accompanied by legal or moral responsibility in a role, liability, or capacity understanding of responsibility that usually attaches to human action.").

[188] Dieter Fleck, *Methods and Means of Combat*, in STEFAN OETER (ed.), THE HANDBOOK OF INTERNATIONAL HUMANITARIAN LAW, 115 (Oxford University Press, 3d edition, 2013) ("The concern to protect the civilian population, as well as combatants, against excessive and exceptionally cruel

immunity to enemy negotiators."[190] The Babylonians, in their *Code of Hammurabi*, also provided for the protection of the oppressed and for the release of hostages upon payment of ransom.[191] As early as 400 BC, Indian laws prohibited certain means of warfare, such as poisoned arrows and the killing of a surrendering enemy.[192] The Greeks and the Romans respected the life of war victims and prisoners of war, respectively.[193] In the Middle Ages, wars were also subject to strict principles, such as St. Augustine's *jus in bello* principles, which provided, inter alia, that war must be fought in a manner to discriminate between proper objects of violence (i.e., combatants) and noncombatants, such as women, children and the elderly.[194] Enlightenment philosophers sought to further entrench the humanization of violence

---

violence might have been the beginning of all moral and philosophical attempts to mitigate the horrors of war.").

[189] Theodor Meron, *The Humanization of Humanitarian Law*, 94(2) THE AMERICAN JOURNAL OF INTENRATIONAL LAW 239, 242 (2000) ("The law of war has always contained rules based on chivalry, humanity, and religious values that were designed to protect noncomatants, especially women, children, and old men, who were presumed incapable of bearing arms and committing acts of hostility. It has also incorporated rules protecting combatants (in matters such as quarter, perfidy, and unnecessary suffering)."); CROWE AND WESTON-SCHEUBER, *supra* note 183, at 5 (these historical principles "illustrate that the formal documents at the heart of modern intenrational humanitarian law reflect a long customary tradition.").

[190] Mary Ellen O'Connell, *Historical Development and Legal Basis*, in OETER, *supra* note 188, at 16.

[191] *Id.*

[192] *Id.*

[193] *Id.*

[194] *Id.*; *Augustine: Political and Social Philosophy*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, *available at* http://www.iep.utm.edu/aug-poso/; Fleck, *supra* note 188, at 123 ("As O'Connell notes, "[t]he humanitarian principles of Islamic legal culture and the moral theological postulates of medieval scholars already contemplated a principle of distinction between combatants and civilians and called for extensive protection of the civilian population.").

in armed conflicts by advocating for the prohibition of targeting prisoners or war.[195] From this view followed the general principle that "acts of hostility may only be directed against the armed forces of the adversary, not against the civilian population which takes no part in the hostilities."[196]

The United States played an important role in the codification of these principles. During the Revolutionary War, George Washington agreed with the British that the conflict would be "carried on agreeable to the rules which humanity is formed" and "to prevent or punish every breach of the rules of war within the sphere of [their] respective commands."[197] During the American Civil War, President Abraham Lincoln approved a set of instructions to the Union Army, entitled "Instructions for the Government of Armies of the United States in the Field."[198] These instructions, known as the Lieber Code, set out a set of rules to govern the armies' conduct, including a prohibition against committing violence against

---

[195] *See e.g.* JEAN-JACQUES ROUSSEAU, ON THE SOCIAL CONTRACT 6 (Dover Thrift Edition, 2003) ("as soon as [soldiers] lay them down and surrender they become once more merely men, whose life no one has any right to take").

[196] O'Connell, *supra* note 190, at 19-20.

[197] *Law of War Manual*, U.S. DEPARTMENT OF DEFENSE (Jun. 2015) [hereinafter "DoD Law of War Manual"], at ii.

[198] *See* FRANCIS LIEBER, INSTRUCTIONS FOR THE GOVERNMENT OF ARMIES OF THE UNITED STATES IN THE FIELD (D. Van Nostrand, 1863) *available at* https://archive.org/details/governarmies00unitrich [hereinafter "Lieber Code"].

civilians, robbery, and rape.[199] It also provided for hostage rights and prohibited the punishment of prisoners of war.[200]

The Lieber Code inspired other states to implement similar codes and "served as a template for international codifications of the law of war."[201] Indeed, the Code is the "origin of what has come to be known as 'Hague Law'"[202] which sets out the law of armed conflict from the "standpoint of the soldier, in the sense that it takes the form of a statement of the rights and duties of the military in a conflict."[203] The Geneva Conventions (also known as "Geneva Law"),[204] on the other hand, focuses on the law of armed conflict "from the standpoint of the 'victims' of war", including civilians, the wounded, the sick, and prisoners of war."[205] As O'Connell notes, Geneva Law "does not purport to define the rights and duties of the military but rather to lay down certain basic obligations designed to protect those victims, while

---

[199] Lieber Code, *supra* note 198, at ¶ 44 ("All wanton violence committed against persons in the invaded country, all destruction of property not commanded by the authorized officer, all robbery, all pillage or sacking, even after taking a place by main force, all rape, wounding, maiming, or killing of such inhabitants, are prohibited under the penalty of death, or such other severe punishment as may seem adequate for the gravity of the offence.").

[200] *Id.*, at ¶ 56.

[201] DoD Law of War Manual, *supra* note 197, at ii.

[202] O'Connell, *supra* note 190, at 22 ("IHL Is a vast body of substantive rules comprising what are traditionally called 'the law of the Hague' and 'the law of Geneva'. The former set of rules includes some Hague Conventions of 1899 or 1907 on international warfare. These rules, in addition to providing for th various categories of lawful combatants, primarily regulate combat actions (means and methods of warfare) and the treatment of persons who no longer take part in armed hostilities (prisoners of war). The so-called 'law of Geneva' comprises the various Gevena Conventions (at present the four Conventions of 1949 plus the two Additional Protocols of 1977), and is essentially designed to regulate the treatment of persons who do not, or no longer, take part in armed conflict (civilians, the wounded, the sick and shipwreacked, as well as prisoners of war)").

[203] *Id.*, at 22.

[204] ANTONIO CASSESE, INTERNATIONAL CRIMINAL LAW 81 (Oxford University Press, 2008).

[205] O'Connell, *supra* note 190, at 22.

leaving to customary law and Hague Law questions which do not fall within its provisions."[206]

The normative evolution of the international legal order has shifted from an emphasis on state security to a "focus on human security."[207] As the International Criminal Tribunal for the Former Yugoslavia (ICTY) noted in *Prosecutor v. Dusko Tadic*, "[a] State-sovereignty-oriented approach has been gradually supplanted by a human-being-oriented approach. Gradually the maxim of Roman law *hominum causa omne jus constitutum* (all law is created for the benefit of human beings) has gained a firm foothold in the international community as well."[208]

While the anthropocentric focus of the law of armed conflict is to be welcomed,[209] it must not be misinterpreted as a normative justification for the seemingly corollary proposition that violence must only be inflicted by a human being. So if we are to take a gun as an example, it would preposterous not to acknowledge that, in functional terms, the gun and bullet share practical responsibilities for the infliction of violence. While moral and legal responsibility lay exclusively with the human subject, practical responsibilities can be distributed across

[206] *Id.*, at at 22.

[207] RUTI G. TEITEL, HUMANITY'S LAW 4 (Oxford University Press, 2011) ("The normative foundations of the international legal order have shifted from an emphasis on state security— that is, security as defined by borders, statehood, territory, and so on—to a focus on human security: the security of persons and peoples. In an unstable and insecure world, the law of humanity—a framework that spans the law of war, international human rights law, and international criminal justice—reshapes the discourse of international relations."); Blum, *supra* note 79, at 48 ("I argue that the "humanization of international humanitarian law" marks a shift from collectivism toward cosmopolitan individualism in the regulation of wartime conduct.").

[208] *Prosecutor v. Dusko Tadic*, IT-94-1, Appeals Chamber, Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction (ICTY, 1995), at ¶ 97.

[209] The law can also be used to advance non-human ends, such as protecting animals, the environment, and perhaps one day, artificial intelligence.

a network that combines humans and non-human actants to source, produce, organize, and ultimately deliver violence on a target. Indeed, the "package" in which violence "comes in, machine or human, is not the deepest moral principle."[210] International Humanitarian Law is designed to "regulate the treatment of persons—civilian or military"[211] in armed conflicts, and imposes no requirement that the final causal element in the chain of violence is a human being.[212] To give the law a "human face,"[213] as Theodor Meron put it, does not necessarily require that a human face shadow all aspects of violence. Other forms of human-made intelligence can also advance the humanization of war.

The next two sections will argue that human use of VIS capable of autonomous functions is compatible with international obligations. It will also argue that, far from the case of deodands where no human is responsible,[214] human beings and human institutions remain responsible for the use of VIS. Much like a manufacturer is responsible for the products she makes, humans, as manufacturers of violence, are responsible for its effects.

---

[210] Kenneth Anderson and Matthew Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, HOOVER INSTITUTION, Research Paper, 16 (2013).

[211] O'Connell, *supra* note 190, at 11.

[212] *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare, Measures to Implement Article 36 of Additional Protocol I of 1977*, INTERNATIONAL COMMITTEE OF THE RED CROSS, [hereinafter ICRC Legal Review Guide], 3 (2006) ("IHL consists of the body of rules that apply during armed conflict with the aim ofprotecting persons…").

[213] Meron, *supra* note 189, at 239.

[214] Mind the Gap, *supra* note 66, at 2 ("A commander would nevertheless still escape liability in most cases.").

International humanitarian law governs the means (weapons) and methods (tactics)[215] of warfare. Accordingly, two branches of IHL, namely "weapons law" and "targeting law", govern the development, deployment, and use of VIS.[216] Weapons law governs the legality of the nature and intended use of the weapon. Targeting law governs the use of the weapon in international armed conflicts. The application of these laws to VIS shows that far from *non liquet*, international law is clear and adequate to regulate the development and use of VIS.[217]

Before considering in more detail the rules governing permissible conduct in international armed conflicts, four broad observations should be made. First, IHL and ICL obligations apply to human beings and human institutions, not the weapons themselves.[218] While weapons law dictates the legality of a weapon, it does not impose an obligation *on the weapon* to satisfy those requirements. While a weapon

[215] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts art. 35-36, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Additional Protocol I]; ICRC Legal Review Guide, *supra* note 212, at 932 ("The combatants' right to choose their means and methods of warfare is limited by a number of basic IHL rules regarding the conduct of hostilities, many of which are found in Additional Protocol I of 1977 on the protection of victims of international armed conflicts. Other treaties prohibit or restrict the use of specific weapons such as biological and chemical weapons, incendiary weapons, blinding laser weapons and landmines, among others. In addition, many of the basic rules and specific prohibitions and restrictions on means and methods of warfare may be found in customary international law."); Corn, *supra* note 102, at 209.

[216] Anderson, Reisner, and Waxman, *supra* note 158, at 400-401.

[217] *See* Question from Martin Caton MP, Hansard Citation: HC Deb, 6 March 2013, c1021W (noting the UK Government's position that "the Government considers that the existing provisions of international humanitarian law are sufficient to regulate the use of these weapons and therefore we have no plans to call for an international ban. However, we remain firmly committed to their effective control.").

[218] O'Connell, *supra* note 190, at 38 ("The obligations of a state under international humanitarian law are binding not only upon its government and its supreme military command but also upon every individual.")

may be the object of weapons law, it is not its subject. At all times legal obligations rest with human subjects.[219] As the DoD's *Law of War Manual* notes:

> The law of war rules on conducting attacks (such as the rules relating to discrimination and proportionality) impose obligations on persons. These rules do not impose obligations on the weapons themselves; of course, an inanimate object could not assume an "obligation" in any event. Thus, it is not the case that the law of war requires that a weapon determine whether its target is a military objective . . . . Rather, it is persons who must comply with the law of war. For example, persons may not use inherently indiscriminate weapons.[220]

Second, the humanization of the law is curtailed by military necessity, resulting in an approach to warfare that is "limited"[221] and whose "only legitimate purpose is to weaken the military capacity" of the enemy.[222] This approach "requires every belligerent to strike a balance between the conflicting concerns of humanity and military necessity."[223] Further, Article 35(1) of the Additional Protocol I of the

---

[219] Sassóli, *supra* note 150, at 323 ("Only human beings are subject to legal rules. In the case of autonomous weapons, IHL is addressed to those human beings who devise, produce and program them, as well as those who decide upon their use. I reject the idea that IHL is inadequate to regulate autonomous weapons because they would be situated somewhere between weapon systems and combat- ants, and further reject the suggestion that a new category with new rules should be created to regulate them. The difference between a weapon system and a human being is not quantitative but qualitative; the two are not situated on a sliding scale, but on different levels-subjects and objects. A combatant is a human being, only he or she is an addressee of legal obligations. However far we go into the future and no matter how artificial intelligence will work, there will always be a human being at the starting point. In my understanding, an autonomous weapon system will always operate within the limits of its software; software designed by humans. It is the human being who will decide whether a machine will be created and who will create it. Even if one day robots construct other robots, there will still be the need for a human being to develop the first robot and instruct it as to how to construct new robots. This human being is bound by the law; the machine is not bound by the law.").

[220] DoD Law of War Manual, *supra* note 197, at ¶ 6.5.9.3.

[221] Fleck, *supra* note 188, at 122; ICRC Legal Review Guide, *supra* note 212, at 3 ("IHL sets limits on armed violence in wartime in order to prevent, or at least reduce, suffering.").

[222] *Military necessity*, INTERNATIONAL COMMITTEE OF THE RED CROSS, Glossary, *available at* https://casebook.icrc.org/casebook/doc/glossary/military-necessity-glossary.htm.

[223] Fleck, *supra* note 188, at 122 (To curtail the objective of humanization of warfare is justified only in so far as military necessity inevitably requires a certain military operation, not to mention the further condition that the damage inflicted must be proportionate to the military advantage sought.").

Geneva Convention 1977 states that "[i]n any armed conflict, the right of the Parties to the conflict to choose methods or means of warfare is not unlimited."[224]

Third, new weapons are subject to legal reviews. Article 36 of the Additional Protocol I states that "in the study, development, acquisition or adoption of a new weapon, means or method of warfare, [a State] is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by [IHL] or by any other rule of international law applicable to the [] Party." In making this assessment, the State must evaluate "not only the weapon's design and characteristics (the 'means' of warfare) but also how it is to be used (the 'method' of warfare), bearing in mind that the weapon's effects will result from a combination of its design *and* the manner in which it is to be used."[225]

Finally, we cannot underestimate international law's flexibility and adaptability. Indeed, "[t]hroughout history, IHL has shown a considerable capability to adapt its functional rules to meet challenges presented by newly developed weapon systems. IHL contains general principles and generally applicable rules to a variety of weapon systems, rather than focusing on one individual technology."[226] As this thesis argues in Chapter IV, international law seeks to regulate means and methods of

---

[224] Additional Protocol I, art. 35(1).

[225] ICRC Legal Review Guide, *supra* note 212, at 17; Schmitt and Thurnher, *supra* note 54, at 271 ("While some commentators suggest that a disagreement exists as to whether Article 36 restates customary international law, the obligation to conduct legal reviews of new *means* of warfare before their use is generally considered, and correctly so, reflective of customary international law. Consensus is lacking as to whether an analogous requirement exists to perform legal reviews of new *methods* of warfare.").

[226] Wagner, *supra* note 39, at 16 ("The existing rules of IHL are capable of responding to AWS, despite considerable differences in opinion that exist in interpreting these rules.").

human violence (i.e., violence designed by humans but delivered via many conduits), rather than just regulating violence unleashed by human beings.

Weapons Law

Under weapons law, the legality of a weapon system, like a VIS, depends on five basic rules. First, it is necessary to determine whether there is a specific prohibition on that weapon system under international treaty law (for example, the Convention on Certain Conventional Weapons[227] prohibits blinding laser weapons[228]) and customary international law (which, for example, prohibits the use of poison in weapons).[229] Second, the weapon system cannot, by its nature, be indiscriminate.[230] What this means is that the weapon cannot have been designed for use in an indiscriminate manner. This is to be contrasted with actually using a discriminate weapon in an indiscriminate manner, which is governed by targeting law.[231] Third, the weapon system cannot be "of a nature" to cause "unnecessary suffering or

---

[227] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (CCW), Geneva, 10 October 1980, and Amendment to Article 1, 21 December 2001.

[228] Protocol on Blinding Laser Weapons (Protocol IV to the 1980 Convention), 13 October 1995 [hereinafter Blinding Laser Protocol].

[229] JEAN-MARIE HENCKAERTS AND LOUISE DOSWALD-BECK (eds.), CUSTOMARY INTERNATIONAL HUMANITARIAN LAW (Cambridge University Press, 2005).

[230] Additional Protocol I, art. 51(4)(c) ("Indiscriminate attacks are prohibited. Indiscriminate attacks are: . . . those which employ a method or means of combat the effects of which cannot be limited as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction."); *see also* Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.CJ. 226 (July 8), 78 [hereinafter Nuclear Weapons], at ¶ 78 ("States must never make civilians the object of attack and must consequently never use weapons that are incapable of distinguishing between civilian and military targets").

[231] Anderson, Reisner, and Waxman, *supra* note 158, at 399.

superfluous injury"[232] to combatants.[233] As the ICJ put it in its *Advisory Opinion on Nuclear Weapons*, "'unnecessary suffering' stands for a harm greater than that unavoidable to achieve legitimate military objectives."[234] Fourth, the weapon system can be illegal per se if its effects cannot be "limited" as required by international humanitarian law.[235] An example here would be the use of a biological weapon whose biological agents cannot be controlled.[236] Finally, the weapon must not offend "the dictates of public conscience"[237]—an obligation that is also known as the Martens Clause. The ICJ has interpreted the clause as a "means of addressing the rapid evolution of technology" by preserving "the pre-existing customary law" of IHL.[238] In practical terms, this means that customary international law applies to new means and methods of warfare. Indeed, it is a "failsafe mechanism meant to address

---

[232] Additional Protocol I, art. 35(2) ("It is prohibited to employ weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering.").

[233] Anderson, Reisner, and Waxman, *supra* note 158, at 400.

[234] Fleck, *supra* note 188, at 126.

[235] Additional Protocol I, art. 51(4)(c) ("Indiscriminate attacks are prohibited. Indiscriminate attacks are: . . . those which employ a method or means of combat the effects of which *cannot be limited* as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction.") (emphasis added).

[236] Anderson, Reisner, and Waxman, *supra* note 158, at 400 ("once released, it goes where it goes").

[237] As the ICJ notes, "A modern version of th[e] [Martens] clause is to be found in Article 1, paragraph 2, of Additional Protocol 1 of 1977, which reads as follows: "In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience." Nuclear Weapons, *supra* note 230, at ¶ 78.

[238] Nuclear Weapons, *supra* note 230, at ¶ 78, 84.

lacunae in the law" and "does not act as an overarching principle that must be considered in every case."[239]

Against this backdrop, it is clear that weapons law does not categorically prohibit the use of VIS.[240] First, there is no treaty prohibiting the use of weapons capable of autonomous engagement. Of course, a VIS may incorporate sub-weapon systems that may be prohibited by weapons law. For example, a VIS that uses a blinding laser weapon would contravene international customary law and the 1995 Protocol on Blinding Laser Weapons.[241] It is unlikely, however, that this prohibition would necessarily extend to the VIS technology employing it, unless the two are so connected that they are effectively one and the same weapon system. The question is therefore whether the VIS is *making use* (i.e., as a platform)[242] of the blinding laser, or whether the laser is so integrated with the VIS that the two are functionally one weapon.

---

[239] Schmitt and Thurnher, *supra* note 54, at 275.

[240] Anderson, Reisner, and Waxman, *supra* note 158, at 401 ("None of these rules renders a weapon system illegal per se solely on account of it being autonomous. If a fully autonomous weapon system were supplied with sufficiently reliable parameters and it were able to act on them so as to be able to strike specific targets on the same legal terms of discrimination that would apply to a human soldier, that the weapon system was "autonomous" would not violate the "indiscriminate by nature" rule. Although some might view an autonomous weapon system as "uncontrollable," its effects are not uncontrollable within the meaning of the legal provision."); Schmitt and Thurnher, *supra* note 54, 233 ("The likelihood of an autonomous weapon system being unlawful *per se* is very low.").

[241] Blinding Laser Protocol, *supra* note 228; *see also* HENCKAERTS AND DOSWALD-BECK, *supra* note 229, at 292 r.86 (Cambridge University Press, 2005).

[242] Christof Heyns, *Autonomous weapons systems: living a dignified life and dying a dignified death*, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, 6 ("AWS . . . are weapon platforms, and any weapon can in principle be fitted onto an AWS").

Second, VIS are not by nature indiscriminate. Intelligence, by its very nature, includes the ability to distinguish and organize patterns in separate categories.[243] Advances in robotics and in artificial intelligence (particularly machine learning) will enable the development of VIS that are capable of distinguishing between civilians and combatants.[244] Of course, a particular VIS may be developed with the nefarious purpose of being indiscriminate. But in addition to providing little military advantage and being illegal per se, that weapon would be far from "intelligent." The prohibition in this scenario would attach to a particular type of weapon system geared for that specific purpose, not to intelligent (including autonomous) systems in general. Much will therefore depend on how artificial intelligence, particularly code, is put to use. At bottom, code—as weaponized data—can be used for a variety, if not infinite, purposes. But there is nothing that makes code automatically indiscriminate.[245]

---

[243] FRITZ HENN, NORMAN SARTORIUS, HANFRIED HELMCHEN, AND HANS LAUTER, CONTEMPORARY PSYCHIATRY 8 (Springer, 2001) ("The mechanics of intelligence include elementary processes of information processing such as information input, visual and motor memory, and basic perceptual-cognitive processes such as discrimination, comparison and categorization").

[244] Schmitt and Thurnher, *supra* note 54, at 247 ("Modern sensors can, inter alia, assess the shape and size of objects, determine their speed, identify the type of propulsion being used, determine the material of which they are made, listen to the object and its environs, and intercept associated communications or other electronic emissions. They can also collect additional data on other objects or individuals in the area and, depending on the platform with which they are affiliated, monitor a potential target for extended periods in order to gather information that will enhance the reliability of identification and facilitate target engagement when the risk of collateral damage is low. Even software for autonomous weapon systems that enables visual identification of individuals, thereby enhancing accuracy during autonomous 'personality strikes' against specified persons, is likely to be developed. These and related technological capabilities auger against characterization of autonomous weapon systems as unlawful per se solely based on their autonomous nature.").

[245] Further, there may be instances where it is lawful to use an indiscriminate autonomous weapon that is used in areas populated exclusively by active military personnel. Schmitt and Thurnher, *supra* note 54, at 246 ("Typical examples would include the employment of such systems for an attack on a tank formation in a remote area of the desert or from warships in areas of the high seas far from maritime navigation routes.").

Third, VIS do not "automatically"[246] violate the prohibition on unnecessary suffering and superfluous injury. The prohibition is directed at regulating the effects of a weapon on a targeted individual, "not the manner of engagement."[247] Similar to the blinding laser example above, a VIS could be used as a platform for a weapon or sub-weapon system that violates this prohibition.[248] But the opposite holds just as true. Much will therefore depend on what type of violence the VIS will inflict.

Fourth, VIS, unlike conventional weapons, can be controlled and limited through code. Code will enable human agents to program VIS to comply with targeting law and other international legal obligations. Finally, VIS do not contravene the "dictates of public conscience" because there is nothing in their nature that makes them incapable of complying with IHL. To the contrary, as this thesis will explore below, VIS may be better at complying with IHL than human soldiers.[249] Public conscience may thus require the use of VIS over human soldiers in certain environments.[250]

---

[246] Schmitt and Thurnher, *supra* note 54, at 245.

[247] *Id.*, at 245.

[248] *Id.*, at 233 ("whereas some conceivable autonomous weapon systems might be prohibited as a matter of law, the use of others will be unlawful only when employed in a manner that runs contrary to the law of armed conflict's prescriptive norms governing the "conduct of hostilities.").

[249] *Id.*, at 247 ("It must be emphasized that as a matter of law, more may not be asked of autonomous weapon systems than of human-operated systems. For example, some opponents of autonomous weapons contend a ban is necessary because autonomous weapon systems may be deceived, as in the case of 'concealing weapons or by exploiting their sensual and behavioural limitations,' and thereby have difficulty distinguishing civilians from combatants. Yet, asymmetrically disadvantaged enemies have been feigning civilian or other protected status to avoid being engaged by humano-perated weapon systems for centuries.").

[250] BROOKS, *supra* note 4, at 137 ("Might there be a legal and ethical *obligation* to use 'killer robots' in lieu of 'killer humans'?").

Targeting Law

Under targeting law, the use of lawful weapons must adhere to two fundamental rules: distinction and proportionality. There are also related considerations, such as the requirement to take all feasible precautions in an attack and obligations on how to resolve doubts about the status of persons in international armed conflicts.

Distinction.

IHL requires parties to a conflict to "at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[251] This "intransgressible"[252] principle[253] is further established in Articles 51(2) and 52(2) of Additional Protocol I:

> Article 51(2). The civilian population as such, as well as individual civilians, shall not be the object of attack. […]
>
> Article 52(2). Attacks shall be limited strictly to military objectives. […][254]

---

[251] Additional Protocol I, art. 48.

[252] Nuclear Weapons, *supra* note 230, at ¶ 79.

[253] Wagner, *supra* note 39, at 20 (noting that some have argued that this interpretation elevates the principle of distinction to *jus cogens* and arguing that there is also "little doubt" that the principle "has attained the status of customary international law . . . . The Court appears to have elevated the principle of distinction to the level of *jus cogens* when it considered it to "constitute [an] intransgressible principle . . . of international customary law.").

[254] *See also* Wagner, *supra* note 39, at 18 ("Distinguishing between a person and an object that possesses a military character—as opposed to an object that is of a civilian character—is the first step in deciding whether a person or object can be lawfully targeted. IHL is based on the assumption that an individual who is not a combatant is a civilian. This assumption is incorporated in the 1868 Declaration of St. Petersburg, the earliest international instrument in the field of international humanitarian law.")

In other words, what the principle of distinction means in practice is that "the *aiming point* for the use of military force must be a military target,"[255] and State must therefore "never make civilians the object of attack."[256]

Critics of autonomous weapon systems argue that computer systems will find it challenging, if not impossible, to satisfy this criteria because distinction involves qualitative decision-making and context-dependent analyses that only a human is able to make.[257] For example, "[n]ot only would [the weapon] have to be able to distinguish civilians from military personnel, but it must also decide if a civilian is taking a direct part in hostilities."[258] Other difficulties may include "distinguishing the agonized face of a person in fear for her or his children from a threatening face" and "distinguishing children playing from threats."[259] One commentator notes the need "to be able to specify every element [of combat] in sufficient detail for a

---

[255] ARMIN KRISHNAN, KILLER ROBOTS: LEGALITY AND ETHICALITY OF AUTONOMOUS WEAPONS 93 (Routledge, 2009).

[256] Nuclear Weapons, *supra* note 230, at 78.

[257] Wagner, *supra* note 39, at 23 ("This type of analysis does not rely on quantitative data—as is the case with the distinction between a tank and a school bus—but rather requires qualitative analysis."); *see also The UK Approach to Unmanned Aircraft Systems*, Joint Doctrine 2/11, MINISTRY OF DEFENSE (Mar. 2011) [hereinafter UK Joint Doctrine], at ¶ 507 ("From this position, it would be only a small technical step to enable an unmanned aircraft to fire a weapon based solely on its own sensors, or shared information, and without recourse to higher, human authority. Provided it could be shown that the controlling system appropriately assessed the [international humanitarian law] principles (military necessity; humanity; distinction and proportionality) and that [rules of engagement] were satisfied, this would be entirely legal . . . . In practice, such operations would present a considerable technological challenge and the software testing and certification for such a system would be extremely expensive as well as time consuming. Meeting the requirement for proportionality and distinction would be particularly problematic, as both of these areas are likely to contain elements of ambiguity requiring sophisticated judgement. Such problems are particularly difficult for a machine to solve and would likely require some form of artificial intelligence to be successful. ").

[258] Wagner, *supra* note 39,at 22.

[259] Wagner, *supra* note 39,at 23.

computer to be able to operate on it."[260] In other words, critics argue that these situations are not "easily programmable."[261]

While these are legitimate concerns, they underestimate the power of artificial intelligence ("AI"). There are two approaches to AI. The first is known as "knowledge engineering,"[262] which requires the programmer to code (or preprogram) all the necessary knowledge into the system's knowledge base. This rationalist[263] approach is the traditional way of thinking of computer programming. *See* Figure 1. But pre-programming knowledge and data into the system is too labor-intensive and failure-prone.[264] Indeed, it would be a very difficult task to preprogram all the possible "eventualities"[265] in complex war environments.

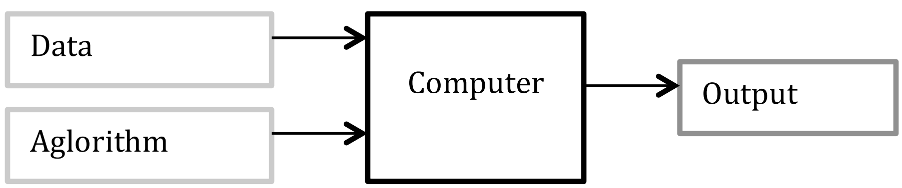

Figure 1. Traditional Programming Approach[266]

[260] Noel E. Sharkey, *The evitability of autonomous robot warfare,* 94 INTERNATIOANL REVIEW OF THE RED CROSS, 886, at 789 (2012).

[261] Wagner, *supra* note 39, at 22; Marchant et al., *supra* note 32, at 284 ("Furthermore, increasing complexity may lead to *emergent behaviors*, i.e., behaviors not programmed but arising out of sheer complexity.").

[262] DOMINGOS, *supra* note 18, at 34 ("Knowledge engineering: knowledge can't be learned automatically; it must be programmed into the computer by humans.")

[263] *Id.*, at 57 ("The rationalist likes to plan everything in advance before making the first move.").

[264] *Id.*, at 90.

[265] *Id.*, at 280 ("There is no universal answer and no way to program a computer with all the eventualities.").

[266] Pedro Domingos, *The Secrets of Machine Learning Revealed*, UNIVERSITY OF WASHINGTON PRESENTATION, at 5 (on file with author).

The second approach, known as machine learning, provides a more attractive—and surprisingly largely ignored—alternative in this context.[267] *See* Figure 2. Machine learning is a form of artificial intelligence that is able to learn from experience.[268] By learning from experience, a system is able to change its "behavior in a way that makes [it] perform better in the future."[269] Thus, in contrast to pre-programming a VIS with all potential possibilities, machine learning will enable it to teach itself lawful courses of action based on algorithmic parameters, modeling, and real-time data.[270] Unlike the rationalist approach of knowledge engineering, machine learning is empirical in the sense that it "prefers to try things and see how they turn out."[271] In effect, it enables the programs to "program themselves."[272]

---

[267] DOMINGOS, *supra* note 18, at 280.

[268] TOM M. MITCHELL, MACHINE LEARNING 2 (McGraw-Hill, 1997) ("A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E").

[269] IAN H.WITTEN, AND EIBE FRANK, DATA MINING: PRACTICAL MACHINE LEARNING AND TECHNIQUES WITH JAVA IMPLEMENTATIONS 6 (Morgan Kaufmann, 2006) (defining data mining).

[270] DOMINGOS, *supra* note 18, at 74 ("learning is a race between the amount of data you have and the number of hypotheses you consider. More data exponentially reduces the number of hypotheses that survive").

[271] DOMINGOS, *supra* note 18, at 57.

[272] *Id.*, at xi.

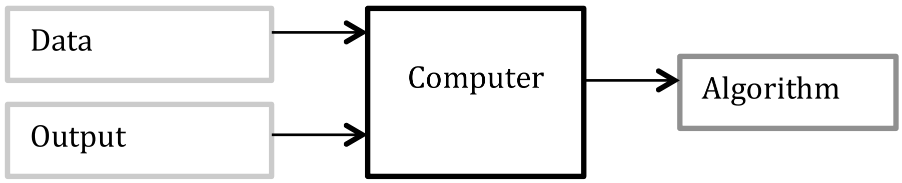

Figure 2. Machine Learning Approach[273]

There are three broad approaches to machine learning: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning refers to "working with a set of labeled training data."[274] As Sugiyama notes, "[t]he objective of supervised learning is to acquire the *generalization ability*, which refers to the capability that an appropriate answer can be guessed for unlearned [scenarios, . . . which means that] the user does not have to teach everything to the computer, but the computer can automatically cope with unknown situations by learning only a fraction of knowledge."[275] Examples of applications that use supervised learning include "hand-written letter recognition, speech recognition and image recognition."[276] Unsupervised learning is where the programmer "let[s] the algorithm find a hidden pattern in a load of data."[277] Here the system will autonomously collect data, or organize apparently random data into clusters to make sense of the data and extract useful knowledge from it.[278] Applications include "system diagnosis, security, event

[273] Pedro Domingos, *The Secrets of Machine Learning Revealed*, University of Washington Presentation, at 5.

[274] JASON BELL, MACHINE LEARNING 3 (Wiley, 2015).

[275] MASAHI SUGIYAMA, INTRODUCTION TO STATISTICAL MACHINE LEARNING 3 (Morgan Kaufman, 2016).

[276] SUGIYAMA, *supra* note 275, at 3.

[277] BELL, *supra* note 274, at 3.

[278] *Id.*, at 3.

detection, and social network analysis."[279] Finally, reinforcement learning aims to achieve the generalization ability without the user providing data. With this approach, a human "*evaluates* the [computer's] behavior and gives feedback about it [and . . .] based on [that] feedback, . . . the [computer] improve[s] [its] behavior to maximize the supervisor's evaluation."[280] This approach is "an important model of the behavior of humans and robots" and it has been applied "to various areas such as autonomous robot control"[281] and would thus be particularly suitable for the development of VIS.[282]

Researchers are seeking to perfect these techniques in many ways. There are currently five schools of thought focused on developing these technologies: symbolists, connectionists, evolutionaries, bayesians, and analogizers. Symbolists believe that "all intelligence can be reduced to manipulating symbols."[283] For example, a grossly simplified rules-based approach to the principle of distinction

---

[279] *Id.*, at 3.

[280] *Id.*, at 3.

[281] *Id.*, at 3.; *see also* Alonso, *supra* note 151, at 236 ("In such scenarios an agent exists in an environment described by a set of possible states. Each time an agent executes an action in a state it receives a numerical reward that indictates the immediate value of this state-action transition – how "good" it is. This produces a sequence of states, actions, and rewards. The agent's task is to learn a policy that maximizes the expected sum of rewards, typically with future rewards discounted exponentially by their delay. . . . This method has been successfully applied to several organizational problems in robotics, control, operations research, games, human-computer interaction, economics-finance, complex simulation, and marketing.").

[282] Guillaume Lample and Devendra Singh Chaplot, *Playing FPS Games with Deep Reinforcement Learning*, ARXIV (Sep. 2016) *available at* https://arxiv.org/abs/1609.05521(finding that reinforcement learning and neural networks can outperform human agents in playing virtual killing games); *id.* at 1 ("Deep reinforcement learning has proved to be very successful in mastering human-level control policies in a wide variety of tasks such as object recognition with visual attention, high-dimensional robot control and solving physics-based control problems. In particular, Deep Qnetworks (DQN) are shown to be effective in playing Atari 2600 games and more recently, in defeating world-class Go players.") (internal citations omitted).

[283] DOMINGOS, *supra* note 18, at 89.

would be a general rule that states: "*If a combatant, then threaten to engage; If not a combatant, then do not threaten to engage.*" This general rule would be dependent on specific sub-rules. For starters, the general rule does not define combatant or civilians.[284] A specific sub-rule that encodes this type of knowledge can be illustrated in the simplified decision tree[285] as shown in Figure 3.

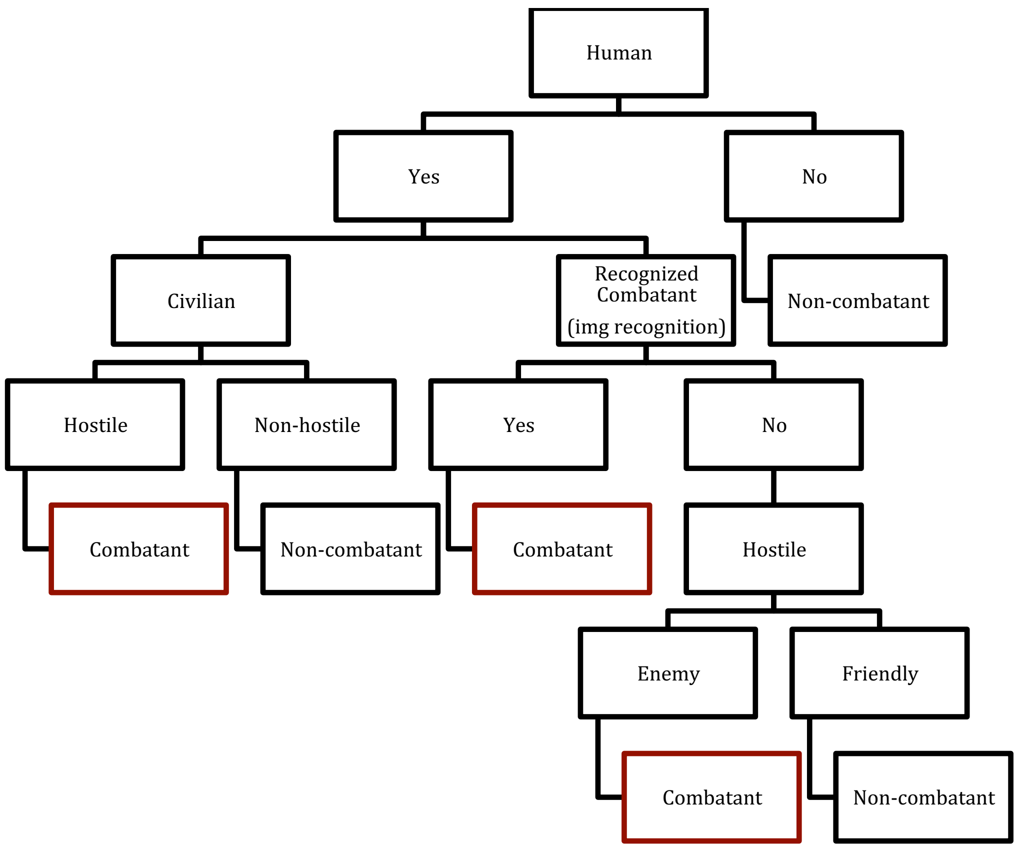

Figure 3. Decision-Tree of Combatants Class

A set of rules would need to be developed for what constitutes a "human," "hostile behavior" (or taking part in hostilities), and what constitutes a "civilian." If

[284] *See* Additional Protocol I, art. 50 (defining civilian for purposes of the Geneva Conventions).

[285] DOMINGOS, *supra* note 18, at 85 ("[T]he symbolist aglorithm of choice is decision tree induction . . . . [For example], Microsoft's Kinect uses decision trees to figure out where various parts of your body are from the output of its depth camera; it can then use their motions to control the Xbox game.").

knowledge is not readily available, the system would be able to fill in gaps in knowledge by using inverse deduction.[286] While symbolism is easy to understand and explain, it has been criticized because "real concepts can seldom be concisely defined by a set of rules . . . . They require weighing and accumulating weak evidence until a clear picture emerges."[287]

To this end, connectionists seek to "reverse engineer the brain"[288] so that the system can learn how to adjust its output with the desire one, a process known as backpropagation. This approach involves creating and interconnecting artificial neurons to develop a neural network. Also known as "deep learning," this approach is already being used by many companies around the world. A simple example is Google's Cat Network, which trained itself to recognize cats by "looking" at 10 million images in youtube videos. Remarkably, Google engineers did not teach the network the concept of a "cat." The network taught itself. As Andrew Ng put it, "[t]he idea is that instead of having teams of researchers trying to find out how to find edges, you instead throw a ton of data at the algorithm and you let the data speak and have the software automatically learn from the data."[289] In this context, research shows that "it is possible to train a face detector without having to label images as

---

[286] *Id.*, at 52 ("Their master aglorithm is inverse deduction, which figures out what knowledge is missing in order to make a deduction go through, and then makes it general as possible.").

[287] DOMINGOS, *supra* note 18, at 91.

[288] *Id.*, at xvii.

[289] Liat Clark, *Google's Artificial Brain Learns to Find Cat Videos*, WIRED (Jun. 26, 2012) *available at* https://www.wired.com/2012/06/google-x-neural-network/.

containing a face or not."[290] Objective classification techniques also allow a system to learn how to recognize and classify objects "without having been explicitly taught the notion of objects."[291] As neuroscience improves our understanding of the brain's neural networks, connectionism will further improve these techniques and enable systems to teach themselves concepts (e.g., who is a civilian? What is a hospital?) in the absence of precise human definitions.

Evolutionaries simulate evolution in artificial systems[292] to create a "brain" that can fine-tune outputs. This approach draws heavily from genetic programming which seeks to get "a computer to do what needs to be done, without telling it how to do it."[293] To achieve this, genetic programming "breeds" a population of programs using Darwinian natural selection and "biologically inspired operations," including

---

[290] Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, Andrew Y. Ng, *Building High-level Features Using Large Scale Unsupervised Learning*, RESEARCH GOOGLE (2012), *available at* http://static.googleusercontent.com/media/research.google.com/en//archive/unsupervised_icml2012.pdf.

[291] Ray Kurzweil, *MIT deep-learning system autonomously learns to identify objects*, KURZWEIL ACCELERATING INTELLIGENCE (May 14, 2016) *available at* http://www.kurzweilai.net/mit-deep-learning-system-autonomously-learns-to-identify-objects (finding that "the same [neural] network can perform both scene recognition and object localization in a single forward-pass, without ever having been explicitly taught the notion of objects"); Richard Socher, Brody Huval, Bharath Bhat, Christopher D. Manning, Andrew Y. Ng, *Convolutional-Recursive Deep Learning for 3D Object Classification*, NEURAL INFORMATION PROCESSING SYSTEMS FOUNDATION (2012) *available at* http://nlp.stanford.edu/pubs/SocherHuvalBhatManningNg_NIPS2012.pdf; Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, NEURAL INFORMATION PROCESSING SYSTEMS FOUNDATION (2012) available at https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf; Andreas Eitel, Jost Tobias, Springenberg, Luciano Spinello, Martin Riedmiller, and Wolfram Burgard, *Multimodal Deep Learning for Robust RGB-D Object Recognition*, ARXIV (Aug. 18, 2015) *available at* http://www2.informatik.uni-freiburg.de/~spinello/eitelIROS15.pdf.

[292] DOMINGOS, *supra* note 18, at xvii.

[293] John R. Koza, *Genetic Programming*, Genetic Programming Website, *available at* http://geneticprogramming.com/tutorial/ [Last accessed: October 5, 2016].

reproduction,[294] crossover,[295] mutation,[296] and architecture-altering operations.[297] After the process is complete, the "best program" in the population is "harvested and designated" as a solution to the problem at hand.[298] Of course, while emulating humans may be desirable in some contexts, the architecture of the human brain may have many faults that we may want to avoid.[299]

Bayesians believe "learning is a form of probabilistic inference"[300] as knowledge is uncertain. At the heart of Bayesian thought is the notion that one's degree of belief in a hypothesis changes as new evidence is received. As Domingos explains, "as you see more data, some models become more likely and some less, until ideally one model stands out as the clear winner."[301] A *simplistic* example can illustrate how this could work. Let us start with a basic hypothesis that a civilian typically wears civilian clothes.[302] Next, let us assume that a VIS is sent to engage the enemy in a military base that is known to be populated exclusively by combatants.

---

[294] *Id.* ("Copy the selected individual program to the new population.").

[295] *Id.* ("Create new offspring program(s) for the new population by recombining randomly chosen parts from two selected programs.").

[296] *Id.* ("Create one new offspring program for the new population by randomly mutating a randomly chosen part of one selected program.").

[297] *Id.* ("Choose an architecture-altering operation from the available repertoire of such operations and create one new offspring program for the new population by applying the chosen architecture-altering operation to one selected program.").

[298] *Id.*

[299] DOMINGOS, *supra* note 18, at 141.

[300] *Id.*, at xvii.

[301] *Id.*, at 144.

[302] This basic premise would need to be modelled so that the system can distinguish between civilian clothing and military clothing.

How should a VIS act if it encounters a human being wearing civilian clothes? Should it engage? The VIS could take certain courses of action to invite reactions from the human, such as giving warnings, or displaying some level of hostility, and based on their reaction, update its hypothesis of the human's status. In mathematical notation, we could say that *P(civilian | hon-hostile behavior*[303]*)*, the conditional probability of a human is a civilian based on behavior, is greater than *P(civilian | civilian clothing)*, its conditional probability given that the human is wearing civilian clothing.[304] Depending on the human's behavior, the hypotheses would be updated to inform the VIS' next course of action.

Finally, analogizers see learning as "recognizing similarities between situations and inferring other similarities."[305] While the challenge here is figuring the "similarity" between two classes of data, the ability to analogize is not foreign to human behavior. From the metaphors we use to how we think, the ability to compare and contrast information is a cornerstone of human intelligence.[306] Powering VIS with analogical reasoning will help it better adjust to unanticipated data[307] as well as learn from "behavior" that adheres to international obligations.

---

[303] For example, not engaging in hostile behavior by producing a weapon in a hostile manner or taking some other hostile action against the VIS or other civilians.

[304] This example assumes the prior probability that humans wearing civilian clothing are less likely to be combatants. Of course, the use of this prior probability depends on the particular combat environment and mission objectives. It may be that it would not be used in missions where intelligence suggests the enemy is deliberately hiding within the civilian population.

[305] DOMINGOS, *supra* note 18, at 53.

[306] Gentner, d., Bowdle, B., Wolff, P., and Boronat, C., *Metaphor is like analogy*, in GENTNER D., HOLOYOAK, K.J., AND KOKINOV, B..N. (eds.), THE ANALOGICAL MIND: PERSPECTIVES FROM COGNITIVE SCIENCE 199 (MIT Press, 2001) ("Metaphor is pervasive in language and thought").

[307] For example, a VIS could use analogical reasoning to make sense of new environments or makeshift weaponry.

The combination of these five paradigms could help develop highly intelligent systems capable of lawful conduct. In *The Master Algorithm*,[308] Pedro Domingos offers a three-step approach on how machine learning can be used to develop VIS. First, we must teach the VIS to recognize relevant data and concepts, such as "data sets of situations where civilians were and were not spared."[309] Second, we need to feed the VIS with a "code of conduct in the form of rules involving these concepts."[310] Third, we must "let the [VIS] learn how to apply the code by observing humans."[311] As Domingos notes, "[b]y generalizing from these examples, the robot can learn an end-to-end model of ethical decision making."[312] Once the VIS' decisions match with a human's "as often as one human agrees with another, the training is complete."[313] Further, VIS development will be subject to rigorous engineering processes such as verification ("did we build the system right?"), validity ("did we build the right system?"), security ("did we build a system sufficiently robust from manipulation?"), and control ("if anything goes wrong, is there an abort

---

[308] In addition to providing an overview of the main approaches to machine learning, Domingos also argues that there may be a master algorithm that combines all schools of thought. DOMINGOS *supra* note 18, at xviii ("[w]hat we really want is a single aglorithm combining the key features of all of them: the ultimate master aglorithm . . . . The Master Aglorithm is to machine learning what the Standard Model is to particle physics or the Central Dogma to molecular biology: a unified theory that makes sense of everything we know to date"); *see also* Boer Deng, *Machine ethics: The robots dilemma*, NATURE (Jul. 1, 2015) ("The challenge — still unresolved — is to combine the [knowledge engineering and machine learning] approaches in a workable way.").

[309] DOMINGOS, *supra* note 18, at 280.

[310] *Id.*, at 280.

[311] *Id.*, at 280.

[312] *Id.*, at 280.

[313] *Id.*, at 280.

mode?").[314]

Machine learning should not be misunderstood as a facile solution to sets of extremely complex problems. But its advantages are too compelling to be ignored in this debate. Human soldiers are subject to fear, anger, fatigue, prejudice, and sexual desire— human emotions and predispositions that can lead to underperformance and violations of international law.[315] Machine learning could enable VIS to learn from humans the context-specific and qualitative assessments involved in combat situations, while allowing for more sophisticated and accurate evaluation of information. For example, VIS could make use of image recognition technology and analogical reasoning in a way that is free from prejudice and bias.[316] Assuming bias

[314] Stuart Russell, Daniel Dewey, Max Tegmark, *Research Priorities for Robust and Beneficial Artificial Intelligence*, FUTURE OF LIFE, Winter 2015, *available at* http://futureoflife.org/data/documents/research_priorities.pdf at 107-108. Further, the use of VIS will be incremental. At first, it could be that VIS accompany soldiers in specific terrains to "learn" from them. Then, VIS can join missions to provide non-lethal support, such as advising human soldiers on compliance with IHL. Finally, once confidence in the system is sufficiently high to adhere to IHL and other applicable rules, VIS can be deployed without real-time human supervision to engage human targets if necessary.

[315] *See* Ronald Arkin, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS 29-32 (Chapman and Hall, 2009) ("Some of the ethical advantages AWS will have over human soldiers may include: the ability to act more conservatively than their human counterparts due to the autonomous system's lack of motivation for self- preservation; the programming to behave in a self-sacrificing manner, if necessary, due to the absence of fear of death; the ability to act without emotions, making autonomous systems insusceptible to anger or fear on the battlefield and therefore able to exercise clearer judgment than humans; and immunity from psychological "scenario fulfillment," which occurs "where humans use new incoming information in ways that only fit their pre-existing belief patterns," and can result in "distortion or neglect of contradictory information in stressful situations.").

[316] *See also The Chip That's Bad at Math*, 119 MIT TECHNOLOGY REVIEW 4, 22 (2016) (reporting on DARPA's research into computer chips that are deliberately programmed to perform mathematical calculations incorrectly which may help to make sense of fuzzy data such as video or other "messy real-world data."); Devin Coldewey, *Google researchers aim to prevent Ais from discriminating*, TECHCRUNCH (Oct. 7, 2016) *available at* https://techcrunch.com/2016/10/07/google-aims-to-prevent-discriminatory-ai-with-equality-of-opportunity-method/?ncid=rss&utm_source=feedburner&utm_medium=feed.&utm_campaign=Feed%3A+Techcrunch+%28TechCrunch%29&utm_content=FaceBook&sr_share=facebook.

and prejudice are not coded into the system,[317] VIS could avoid psychological "scenario fulfillment," which occurs "where humans use new incoming information in ways that only fit their pre-existing belief patterns," and can result in "distortion or neglect of contradictory information in stressful situations."[318] As Rosa Brooks put it, "computers often seem to have *better* judgment" than humans.[319] Brooks notes, for instance, how computers are better than humans at "distinguishing between genuine and faked expressions of pain" and "Google's driverless cars fare better at avoiding accidents than cars controlled by humans."[320]

In addition, VIS will be able to tap into cloud robotics to make use of larger data sets, models, and real-time data to make better-informed decisions (such as distinguishing between types of weapons and cultural objects, for instance),[321] which would arguably far surpass the intelligence of any human soldier.[322] Machines can

---

[317] *See e.g.* Aimee Rawlins, *Math is racist: How data is driving inequality*, CNN MONEY (Sep. 6, 2016), *available at* http://money.cnn.com/2016/09/06/technology/weapons-of-math-destruction/index.html ("From targeted advertising and insurance to education and policing, O'Neil looks at how algorithms and big data are targeting the poor, reinforcing racism and amplifying inequality.").

[318] Ronald Arkin, *The Case for Ethical Autonomy in Unmanned Systems*, in Timothy J. Demy and George R. Lucas (eds.), MILITARY ETHICS AND EMERGING TECHNOLOGIES 76 (Routledge, 2014).

[319] BROOKS, *supra* note 4, at 137.

[320] *Id.*, at 137.

[321] MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGGY AND THE THREAT OF A JOBLESS FUTURE 21 (Basic Books, 2015) ("If one robot employs centralized machine intelligence to learn and adapt to its environment, then that newly acquired knowledge could become instantly available to any other machines accessing the system—making it easy to scale machine learning across large number of robots . . . . The impact of cloud robotics may be most dramatic in areas like visual recognition that require access to vast databases as well as powerful computational capability.").

[322] *Singularity: Ubiquity Interviews Ray Kurzweil*, UBIQUITY (Jan. 2006) *available at* http://ubiquity.acm.org/article.cfm?id=1117663 ("it will be a very powerful combination to combine the subtle and supple powers of human pattern recognition with ways in which machines are already superior.").

also share knowledge faster than human language.[323] Of course, the success and utility of cloud capabilities will depend on the quality of the intelligence on the cloud. But as surveillance technologies and data analysis techniques become increasingly sophisticated,[324] the ability to capitalize on both real-time and archived intelligence will empower VIS to make more informed decisions, particularly in the "fog of war."[325]

Advances in robotics could make VIS extremely agile[326] and situationally aware. For example, improvements in "belief space"[327] are allowing robotics to

---

[323] *Id.* ("Machines can think more quickly than we can. They are much better at logical thinking and much better at remembering things: a $1000 notebook computer can remember billions of things accurately whereas we're hard-pressed to remember a handful of phone numbers. And most importantly, machines can share their knowledge their skills, and their insights at electronic speed, which is a million times faster than human language.").

[324] *See e.g.* DARPA's Fast Lightweight Autonomy Program. Patrick Tucker, *The Military Wants Smarter Insect Spy Drones*, DEFENSE ONE (Dec. 23, 2014) *available at* http://www.defenseone.com/technology/2014/12/military-wants-smarter-insect-spy-drones/101970/ ("The Fast Lightweight Autonomy program, the agency said, "focuses on creating a new class of algorithms to enable small, unmanned aerial vehicles to quickly navigate a labyrinth of rooms, stairways and corridors or other obstacle-filled environments without a remote pilot."); Michael V. Hayden, *To Keep America Safe, Embrace Drone Warfare*, NEW YORK TIMES (Feb. 19, 2016) *available at* http://www.nytimes.com/2016/02/21/opinion/sunday/drone-warfare-precise-effective-imperfect.html?_r=0 ("TARGETED killing using drones has become part of the American way of war. To do it legally and effectively requires detailed and accurate intelligence . . . . And unmanned aerial vehicles carrying precision weapons and guided by powerful intelligence offer a proportional and discriminating response when response is necessary.").

[325] Arkin, *supra* note 318, at 75 ("In the fog of war it is hard enough for a human to be able to effectively discriminate whether or not a target is legitimate. Fortunately, it may be anticipated . . . that in the future autonomous robots may be able to perform better than humans under these conditions").

[326] *See* Boston Dynamics Website, *available at* https://www.youtube.com/user/BostonDynamics [Last accessed October 4, 2016].

[327] ALEC ROSS, THE INDUSTRIES OF THE FUTURE 23-24 (Simon & Schuster, 2016) (Belief space means the "mathematical framework that allows us to model a given environment statistically and develop probabilistic outcomes . . . . [and allows for] the application of algorithms to make sense of new or messy contexts."); *see also* ERIK BRYNJOLFSSON AND ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 52 (W. W. Norton & Company, 2016) (noting imrpovements in SLAM ('simultaneous localization and mapping')); *but see id.*, at 28-29 (noting Moravec's paradox, which is "the discovery by artificial intelligence and robotics researchers that, contrary to traditional assumptions, high-level reasoning

execute tasks once thought the "exclusive domain of humans," such as tasks requiring "situation awareness, spatial reasoning and dexterity, contextual understanding, and human judgment."[328] As Alex Ross notes in *The Industries of the Future*:

> Linked to the cloud, robots can access vast troves of data and shared experience to enhance the understanding of their own belief space. Before being hooked to the cloud, robots had access to very limited data—either their own experience or that of a narrow cluster of robots. They were stand-alone pieces of electronics with capabilities that were limited to the hardware and software inside the unit. But by becoming networked devices, constantly connected to the cloud, robots can now incorporate the experiences of every other robot of their kind, "learning" at an accelerating rate.[329]

Of course, there is the risk that network-connected VIS will be exposed to hacking[330] and tampering.[331] But there are ways to ensure that military-grade communications remain secure. Indeed, in addition to existing security measures, such as DARPA's hack-proof High Assurance Cyber Military Systems,[332] quantum communications may offer a way to develop hack-proof communications between humans and VIS.[333]

---

requires very little computation, but low-level sensorimotor skills require enormous computational resources.").

[328] *Id.*, at 23-24; Fergus Walsh, *Robot operates inside eye in world first*, BBC NEWS (Sep. 9, 2016) *available at* http://www.bbc.com/news/health-37246995 ("Surgeons have used a robot to operate inside the eye and restore sight—in a world first . . . . Surgeons hope the procedure will pave the way for more complex eye surgery than is currently possible with the human hand.").

[329] ROSS, *supra* note 237, at 23-24.

[330] RAY KURZWEIL, THE SINGULARITY IS NEAR 333 (Penguin, 2006) ("An obvious top priority is to develop technology capable of maintaining integrity of communication and preventing either eavesdropping or manipulation of information by hostile forces.").

[331] Schmitt and Thurnher, *supra* note 54, at 242 ("The one real risk is tampering by the enemy or non-State actors such as hackers.").

[332] Kris Osborn, *DARPA Unveils Hack-Proof Drone*, DEFENSE TECH (Oct. 3, 2016) *available at* http://www.defensetech.org/2014/05/21/darpa-unveils-hack-proof-drone/.

[333] *China Launches World's First Quantum Satellite For Hack-Proof Program*, REUTERS (Aug. 16, 2016) *available at* http://www.reuters.com/article/us-china-space-satellite-idUSKCN10R07J (noting in the context of quantum satellites, that "[q]uantum communication boasts ultra-high security

What all these developments mean is that the notion of a VIS as an automaton that must have every eventuality pre-programmed is outdated and does not reflect the potential of current AI research. Machine learning and improvements in robotics offer real possibilities to develop technologies that perform just as well as—if not better than—humans.[334] Indeed, algorithms have been found in many contexts to outperform human "experts." Many studies have demonstrated that "human decision makers are inferior to a prediction formula even when they are given the score suggested by the formula."[335]

The synergistic capabilities of technologies capable of developing highly intelligent systems powered by large data sets about everything from types of weapons to different languages and environments, equipped with sophisticated sensors, such as facial recognition and heat sensing technology, and connected to the

---

as a quantum photon can neither be separated nor duplicated," and that "[i]t is hence impossible to wiretap, intercept or crack the information transmitted through it.").

[334] *See* Sassóli, *supra* note 150, at 319 (There is widespread agreement that the ability to use autonomous weapons in compliance with IHL should not be evaluated against a hypothetical ideal, but instead the comparison should be to human beings)

[335] DANIEL KAHNEMAN, THINKING FAST AND SLOW 224-225 (Farrar, Straus and Giroux, 2011) ("Why are experts inferior to agrlorithms? One reason . . . is that experts try to be clever, think outside the box, and consider complex combinations of features in making their predictions. Complexity may work in the odd case, but more often than not it reduces validity. Simple combinations of features are better. Several studies have shown that human decision makers are inferior to a prediction formula even when they are given the score suggested by the formula! They feel that they can overrule the formula because they have additional information about the case, but they are wrong more often than not . . . . Another reason for the inferiority of expert judgment is that humans are incorrigibly inconsistent in making summary judgment of complex information. When asked to evaluate the same information twice, they frequently give different answers. The extent of the inconsistency is often a matter of real concern. Experienced radiologists who evaluate chest X-rays as 'normal' or 'abnormal' contradict themselves 20% of the time when they see the same picture on separate occasions . . . . A review of 41 separate studies of the reliability of judgments made by auditors, pathologists, psychologists, organizational managers, and other professionals suggests that this level of inconsistentcy is typical . . . . Unreliable judgments cannot be valid predictors of anything.").

cloud, are welcome developments to ensure greater precision[336] and discrimination on the battlefield.[337] Given their potential, a pre-emptive ban on VIS is premature at this stage.

Finally, VIS, like drones, may be increasingly used in missions limited in scope. Indeed, the humanization of IHL has been met by a gradual move toward the "individualization of war."[338] As Brooks put it, "we have already begun to 'individualize' warfare, primarily through targeted drone strikes. In the future, this trend will continue: more and more, we will see highly individualized 'attacks' on specific individuals, in addition to—or in place of—impersonal assaults on 'enemy forces'."[339] When compared to more devastating methods of warfare, such as using nuclear weapons or firebombing, "weapons capable of killing only specific individuals seem like a moral advance."[340] As the individualization of war continues, VIS may likely be used to participate in more precise missions rather than larger offensives, thus minimizing operational risks.

---

[336] Heyns, *supra* note 7, at 13 ("humans are not necessarily superior to machines in their ability to distinguish. In some contexts technology can offer increased precision. For example, a soldier who is confronted with a situation where it is not clear whether an unknown person is a combatant or a civilian may out of the instinct of survival shoot immediately, whereas a robot may utilize different tactics to go closer and, only when fired upon, return fire. Robots can thus act "conservatively" and "can shoot second." Moreover, in some cases the powerful sensors and processing powers of LARs can potentially lift the "fog of war" for human soldiers and prevent the kinds of mistakes that often lead to atrocities during armed conflict, and thus save lives.").

[337] *See* DoD Law of War Manual, *supra* note 197, at ¶ 6.5.9.2 ("In fact, in many cases, the use of autonomy could enhance the way law of war principles are implemented in military operations. For example, some munitions have homing functions that enable the user to strike military objectives with greater discrimination and less risk of incidental harm. As another example, some munitions have mechanisms to self-deactivate or to self-destruct, which helps reduce the risk they may pose generally to the civilian population or after the munitions have served their military purpose.").

[338] BROOKS, *supra* note 4, at 133.

[339] *Id.*, at 132.

[340] BROOKS, *supra* note 4, at 132.

Proportionality.

Closely related to the principle of distinction is the rule "that the use of force and the means employed should always be proportionate to the military objective in order to protect civilians."[341] This rule of customary international law, as codified in Articles 51(5)(b) and 57(2)(a)(iii) of Additional Protocol I, prohibits "an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[342] However, without a definition of what counts as "excessive", proportionality is best understood on a case-by-case basis. The greater the military advantage of an engagement, the more the law will "tolerate the expected collateral damage."[343] While these may at first appear hard concepts to define[344] and quantify,[345] advances in machine learning may enable the

[341] KRISHNAN, *supra* note 255, at 92; *see also* Schmitt and Thurnher, *supra* note 54, at 253 ("An important element of the principle of distinction is the rule of proportionality").

[342] *See also* DoD Law of War Manual, *supra* note 197, at ¶ 6.5.9.3 ("In addition, in the situation in which a person is using a weapon that selects and engages targets autonomously, that person must refrain from using that weapon where it is expected to result in incidental harm that is excessive in relation to the concrete and direct military advantage expected to be gained.").

[343] Schmitt and Thurnher, *supra* note 54, at 254.

[344] Schmitt, *supra* note 148, at 20 ("There is no question that autonomous weapon systems could be programmed to perform CDEM-like analyses to determine the likelihood of harm to civilians in the target area. Moreover, these weapon systems would usually be no less likely to generate a reliable result than CDEM since the latter is heavily reliant on scientific algorithms. The more difficult task for the autonomous weapon system would be assessing military advantage. Given the complexity and fluidity of the modern battlespace, it is unlikely in the near future that, despite impressive advances in artificial intelligence, "machines" will be programmable to perform robust assessments of a strike's likely military advantage . . . . Yet, military advantage algorithms could in theory be programmed into autonomous weapon systems.").

[345] Wagner, *supra* note 39, at 27 ("proportionality assessment is almost entirely a qualitative exercise."); Heyns, *supra* note 7, at 14 ("proportionality assessments often involve qualitative rather than quantitative judgements."); *but see* Schmitt and Thurnher, *supra* note 54, at 255 ("Such calculations require consideration of both expected collateral damage and anticipated military advantage. An effective system already exists for determining the likelihood of collateral damage to objects or persons near a target. The "collateral damage estimate methodology" (CDEM) is a

computation of data and powerful algorithms capable of making sense of complex situations on the ground.[346]

The ability to measure proportionality is not exclusive to the human brain. A proportion is, after all, a mathematical concept. Indeed, computer systems are already used to calculate collateral damage[347] in a number of missions. The U.S. Air Force, for instance, uses computer models to estimate the best approach to "ensure maximum effect on a target with minimum civilian casualties."[348] Thus, even context-specific concepts, such as "military advantage," could "in theory be programmed" into a VIS.[349]

---

procedure whereby an attacking force considers such factors as the precision of a weapon, its blast effect, attack tactics, the probability of civilian presence in structures near the target, and the composition of structures to estimate the number of civilian casualties likely to be caused during an attack.").

346 Schmitt and Thurnher, *supra* note 54, at 256 ("'military advantage' algorithms could theoretically be programmed into autonomous weapon systems.").

347 MARCO ROSCINI, CYBER OPERATIONS AND THE USE OF FORCE IN INTERNATIONAL LAW 165 (Oxford University Press, 2014) ("[C]omputers can be used in support of the application of the law of armed conflict . . . e.g. to maintain target data, estimate the best targeting route or weapon, or calculate collateral damage"); Schmitt, *supra* note 148 , at 19 ("A system already exists for determining the likelihood of collateral damage to objects or persons near a target. The "collateral damage estimate methodology" (CDEM) is a procedure whereby an attacking force considers such factors as the precision of a weapon, its blast effect, attack tactics, the probability of civilian presence in structures near the target, and the composition of structures to estimate the number of civilian casualties likely to be caused during an attack . . . . The commander with authority to authorize the attack makes the proportionality determination as part of the attack's approval process . . . . There is no question that autonomous weapon systems could be programmed to perform CDEM-like analyses to determine the likelihood of harm to civilians in the target area. Moreover, these weapon systems would usually be no less likely to generate a reliable result than CDEM since the latter is heavily reliant on scientific algorithms.").

348 *Off Target. The Conduct of the War and Civilian Casualties in Iraq*, HUMAN RIGHTS WATCH at 19 (2003) *available at* http://www.hrw.org/reports/2003/12/11/target.

349 Schmitt, *supra* note 148, at 20 ("For example, the systems could be *pre*-programmed with unacceptable collateral damage thresholds for particular target sets or situations. As an example, an autonomous weapon system could be pre-programmed with a base maximum collateral damage level of X for a tank; a human would have already made the determination that X generally comports with the proportionality rule. Such thresholds would have to be adjustable by human operators based on the military situation at a particular phase in the conflict, in a particular area of operations, and so forth. Of course, determining the appropriate threshold would be a very subjective endeavour. However, as

Why, then, are we betting against the potential of powerful artificial systems capable of processing information to maximize the mission's objectives while minimizing, and hopefully eliminating,[350] collateral damage? A Violent Intelligent System could better "calculate blast and other weapon effects that cause collateral damage," calculations that would be "far too complex for the warfighter to make in real time."[351] Further, drone surveillance (both aerial and lightweight in urban areas) will help the mapping of the battlefield and identification of targets and civilians. Just like a machine was able to learn how to play the game Go by playing the games several thousands of times against itself, and then beat human players at it, a VIS' operating software could model thousands of eventualities based on given data to provide a suggested "attack plan" that maximizes military advantage and minimizes collateral damage.[352]

---

noted by the ICRC commentary to Additional Protocol I, and as acknowledged in *Losing Humanity*, proportionality determinations necessarily involve a "fairly broad margin of judgment" and "must above all be a question of common sense and good faith for military commanders." Because military advantage is such a context specific value, compliance with the rule of proportionality would require that the base maximum collateral damage threshold be very conservative. Algorithms could be then be developed that would permit the autonomous weapon system to refine the base level threshold to account for specified variables it encountered on a mission. As an example, it would be reasonable to allow the system to increase the level of acceptable collateral damage if it identifies a concentration of enemy tanks, as distinct from a single tank. The concentration poses a greater threat and therefore the military advantage of destroying individual tanks making up the concentration is greater than that of destroying the same tanks when they are operating alone. Similarly, it would be reasonable for the system to adjust the level of acceptable collateral damage based on whether a targeted tank is headed towards or away from the battlefront.").

[350] Sharkey, *supra* note 260, at 1 ("[i]t is proposed here that hi-tech nations should set the specific goal of developing weapons that enable zero civilian casualties and facilitate combatant surrender.")

[351] KRISHNAN, *supra* note 255, at 92 (They could "could perform hundreds of these same calculations in real time, increasing the lethality of the engagement while simultaneously reducing the probability of collateral damage.").

[352] A better way to prescribe this in code would be to ensure that VIS can maximize protection of civilians while still achieving military advantage.

Further, a VIS need not shoot to kill.[353] Indeed, VIS may be able to achieve mission objectives without using lethal force.[354] Alternatives such as "immobilisation or disarmament"[355] may become common practice. For example, the DoD is already working on technology to immobilize the enemy. Dubbed "active denial technology," the technology shoots a "beam of radio frequency millimeter waves toward a specific area," effectively paralyzing the target "with no permanent ill effects" and thus giving the military more time to suppress ambushes and identify targets from civilians.[356] This would give the VIS time to obtain further data to ensure that the strike is proportional. A VIS can also shoot-second if it is unable to determine whether potential targets are actually combatants.

Finally, while operating autonomously, VIS should nonetheless be subject to a "kill or abort switch" if human commanders need to suspend the mission.[357] An abort switch does not undermine the system's autonomy or intelligence. Since

---

[353] Ronald Arkin, *Lethal Autonomous Systems and the Plight of the Non-Combatant*, AISB QUARTERLY, No. 137, 3 (Jul. 2013) ("There is no need for a 'shoot first, ask-questions later' approach, but rather a 'first-do-no-harm' strategy can be utilized instead.").

[354] Cesáreo Gutiérrez Espada and María José Cervell Hortal, *Autonomous Weapons Systems, Drones, and International Law*, 2 REVISTA DEL INSTITUTO ESPAÑOL DE ESTUDIOS ESTRATÉGICOS, 4 (2013) ("In the future, they will be able to employ a less lethal force, thereby avoiding unnecessary deaths."); Gabriella Blum, *The Dispensable Lives of Soldiers*, 2(1) JOURNAL OF LEGAL ANALYSIS 69, 69 (proposing the introduction of "a least-harmful-means test, under which an alternative of capture or disabling of the enemy would be preferred to killing whenever feasible.").

[355] Espada and Hortal, *supra* note 354 at 4.

[356] *Active Denial Technology Fact Sheet,* U.S. DEPARTMENT OF DEFENSE, Non-Lethal Weapons Program (May 11, 2016) *available at* http://jnlwp.defense.gov/Press-Room/Fact-Sheets/Article-View-Factsheets/Article/577989/active-denial-technology-fact-sheet/.

[357] *See* Additional Protocol I, art. 57(2)(b) ("an attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated").

commanders can order subordinates and other actants (e.g., drones) to retreat or abort missions, there may be a case for VIS to be subject to real-time communications with commanders. This would be the case anyways if the VIS is up-linking real-time video footage of the mission or if it is using the cloud to maximize its informational capabilities. Thus, if commanders deem a certain target to no longer provide a "military advantage," the "abort" switch should be used to "cancel the attack."[358]

Precautions and Doubt.

Tied to the principle of distinction is the requirement that the attacker take precautions to minimize harm to civilians and civilian objects.[359] This rule requires the attacker to "take all feasible steps to minimize the risk to civilian casualties, to cancel an operation when disproportionate consequences are expected or intelligence reveals that the targeted object is not military and to warn possibly affected civilians, if this does not contravene the purpose of the attack."[360] In the context of VIS, this may include ensuring that the VIS' code is updated with the latest intelligence and upgraded with the most effective capabilities, that the hardware is upgraded and capable of making use of all its sensors,[361] and that an abort or deactivation

---

[358] *See* Additional Protocol, art. 57(2)(b) ("an attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated").

[359] Additional Protocol I, art. 57(1) ("In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.").

[360] Fleck, *supra* note 188, at 183.

[361] Schmitt and Thurnher, *supra* note 54, at 260 ("The requirement to do everything feasible to verify that the target is a military objective would, for example, require full use of onboard or external sensors capable of boosting the reliability of target identification.").

functionality[362] is available.[363] These would all be "feasible" precautions.[364] Human decisions from the design of the software and hardware through deployment thus carry an ongoing responsibility to ensure that a precautionary approach is taken to minimize civilian casualties.[365]

A precautionary approach also extends to doubts about the status of persons or objects[366] on the battlefield. Article 50(1) of the Additional Protocol I provides that "[i]n case of doubt whether a person is a civilian, that person shall be considered to be a civilian."[367] As a VIS will likely not possess the same bias for self-preservation as a

---

[362] DoD Law of War Manual, *supra* note 197, at ¶ 6.5.9.3 ("the obligation on the person using the weapon to take feasible precautions in order to reduce the risk of civilian casualties may be more significant when the person uses weapon systems with more sophisticated autonomous functions. For example, such feasible precautions a person is obligated to take may include monitoring the operation of the weapon system or programming or building mechanisms for the weapon to deactivate automatically after a certain period of time.").

[363] An abort functionality does not necessarily affect the "autonomy" of the system. Would we say that a soldier is any less autonomous because it must comply with a commander's order to abort a mission?

[364] Schmitt and Thurnher, *supra* note 54, at 60 ("The fulcrum of the verification requirement is the term feasible. Feasible has been interpreted as that which is practicable or practically possible, taking into account all circumstances ruling at the time, including humanitarian and military considerations.") (internal citations omitted).

[365] Fleck, *supra* note 188, at 183 ("The application of the principle of precaution may be guaranteed by so-called persons 'in the loop' who overview the attack and are fed with information on the targeted area and object and in the consequence can stop the automated attack, even if there is an interruption of communication with the automated UCAV. However, even without a person 'in the loop', other human decisions taken in advance, depending on their algorithms, sophistication, and reliability, may result in a positive legal evaluation. So far though, the requirements of customary law and Article 57 AP I can only be met by human decision-making; technology has not advanced far enough yet to address, for example, risks to civilians (autonomous systems).").

[366] Additional Protocol I, art. 52(3) ("In case of doubt whether an object which is normally dedicated to civilian purposes, such as a place of worship, a house or other dwelling or a school, is being used to make an effective contribution to military action, it shall be presumed not to be so used.").

[367] Schmitt and Thurnher, *supra* note 54, at 263 ("The presumption applies both to doubt regarding the status of a targeted individual and as to whether a person is to be considered a civilian in making proportionality calculations and taking feasible precautions in attack.").

human soldier,[368] the advantage of a shoot-second approach is clear in this context: not only will VIS be able to better manage doubt in the absence of fear and the fog of war, but it may also adjust doubt thresholds more accurately depending on real-time information.[369] After all, doubt is the state of not knowing something with a degree of certainty. Powered by cloud robotics and light-speed communications, it is not inconceivable that a VIS will be able to resolve doubt better than a human being. A VIS' sensors may also be able to capture information that a human soldier cannot naturally access by using heat-sensing technology, ultra-sound detection, and on-board surveillance technology connected to the VIS.[370]

Other Operational Benefits of VIS

The use of VIS may bring many other advantages. First, a VIS can be programmed *not* to follow unlawful orders. Since weapons law prohibits certain types of weapons, it is arguable that a VIS' software, as a component of the weapon, *must* include code that prohibits engagements that are contrary to IHL. Second, VIS

[368] Schmitt and Thurnher, *supra* note 54, at 264 ("Autonomous weapon systems arguably possess advantages over humans with respect to doubt. As with other unmanned systems, they are not constrained by the notion of self-preservation. Therefore, the systems could, in some conceivable circumstances, be programmed to either hold their fire until being fired upon or essentially sacrifice themselves to "reveal the presence of a combatant.").

[369] *Id.*, at 264 ("Even if values can be attributable to such variables, it will still be necessary to set the doubt threshold at which an autonomous weapon system will refrain from attack. In a sense, doing so will resemble programming autonomous weapon systems to refrain from attack because of the risk of violating the proportionality rule . . . . An autonomous weapon system may also have adjustable doubt thresholds that can be set before launch to account for the circumstances in which it will be employed (for example, for use in an area where enemy forces have been highly active as distinct from one where they have not.").

[370] For example, the VIS can operate as a platform where lightweight surveillance drones are operationalized to gather real-time intelligence in situations where VIS requires further intelligence to proceed with a mission.

could record its operations using video, sound, and other data (e.g., metadata such as location data and number of engagements) to help monitor the system's compliance with IHL.[371] When working with teams of human soldiers, VIS could also monitor and report real-time observations about the mission's compliance with international obligations.[372] Militaries around the world may decide to voluntarily report this data to further increase transparency in the use of the system.[373] Third, the deployment of VIS arguably replaces human beings on the ground. This in turn reduces the human cost of war,[374] which, assuming the VIS is compliant with IHL,[375] furthers the humanization of international armed conflicts. Fourth, lesser human involvement in

---

[371] *See also* Robin Murphy and David Woods, *Beyond Asimov: The Three Laws of Responsible Robotics*, IEEE INTELLIGENT SYSTEMS (Jul-Aug. 2009) *available at* http://ts-si.org/files/IEEEIS_WebExtra-0709.pdf at 17 ("Robots should carry "black boxes" or recorders to show what they were doing when a disturbance occurred, not only for the sake of an accident investigation but also to trace the robots' behavior in context to aid diagnosis and debugging."); *New Technologies and the modern battlefield: Humanitarian Perspectives: Autonomous weapons*, INTERNATIONAL REVIEW OF THE RED CROSS, *available at* https://app.icrc.org/e-briefing/new-tech-modern-battlefield/part-3-autonomous-weapons.html ("they may bring a new potential to improve compliance with humanitarian law on the battlefield.").

[372] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, Part I: Motivation and Philosophy*, GEORGIA INSTITUTE OF TECHNOLOGY, 7 (2007) *available at* http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf ("autonomous systems maybe capable 'of independently and objectively monitoring ethical behaviour in the battlefield by all parties and reporting infractions that might be observed'."); Marchant et al., *supra* note 32, at 280 ("when working in a team of combined human soldiers and autonomous systems as an organic asset, they have the potential capability of independently and objectively monitoring ethical behavior in the battlefield by all parties and reporting infractions that might be observed. This presence alone might possibly lead to a reduction in human ethical infractions").

[373] *See* Karen DeYoung and Greg Miller, *White House releases its count of civilian deaths in counterterrorism operations under Obama*, WASHINGTON POST (Jul. 1, 2016) *available at* https://www.washingtonpost.com/world/national-security/white-house-releases-its-count-of-civilian-deaths-in-counterterrorism-operations-under-obama/2016/07/01/3196aa1e-3fa2-11e6-80bc-d06711fd2125_story.html (noting the U.S.' publication of statistics related to U.S. drone attacks).

[374] *See* DOMINGOS, *supra* note 18, at 279 ("One of the prime uses of robots is to do things that are too dangerous for humans, and fighting wars is about as dangerous as it get. Robots already defuse bombs, and drones allow a platoon to see over the hill.").

[375] The alternative is to assume that a VIS does not comply with IHL – just a like a human soldier may fail to comply with those rules.

the "doing of violence" could mean fewer sufferers of post-conflict psychological injury, such as post-traumatic stress disorder ("PTSD"). Fifth, we must recognize that humans are perfectable—not perfect beings. There is little evidence to support the assumption that humans hold a perfect record in adhering to IHL. To the contrary, there is a clear "gap between the [international] norms and the reality."[376] Indeed, as former U.N. Secretary General Kofi Annan put it:

> Despite the adoption of the various conventions on international humanitarian and human rights law over the past 50 years, hardly a day goes by where we are not presented with evidence of the intimidation, brutalization, torture and killing of helpless civilians in situations of armed conflict . . . . [T]he parties to conflicts have acted with deliberate indifference to those conventions.[377]

Code may be better able to bridge the gap between norms and reality. As former Pentagon official Rosa Brooks put it:

> Computers may be far *better* than human beings at complying with international humanitarian law. After all, we humans are fragile and panicky creatures, easily flustered by the fog of war. Our eyes face only one direction; our ears register only certain frequencies; our brains can process only so much information at a time. Loud noises make us jump, and fear floods our bodies with powerful chemicals that can temporarily distort our perceptions and judgment . . . . Computers, in contrast, are excellent in crisis and combat situations [because] [t]hey don't get mad, they don't get scared, and they don't act out of sentimentality.[378]

---

[376] Meron, *supra* note 189, at 277; *see also* Wagner, *supra* note 39, at 46 (noting a 2006 report by the Surgeon General's Office found that "appropriate ethical behavior among soldiers and marines deployed in Operation Iraqi Freedom and Operation Enduring Freedom was questionable").

[377] *Protection of Civilians in Armed Conflict*, Report of the Secretary-General, U.N. Doc. S/1999/957, ¶ 2; *see also* BROOKS, *supra* note 4, at 136-137 ("If the U.S. conflicts in Iraq and Afghanistan produced a surfeit of dead and mangled civilians, it's not because of killer robots, it's because of fallible human decision making.").

[378] BROOKS, *supra* note 4, at 136.

Noting that humans have "historically been quite bad at distinguishing between combatants and civilians, even in traditional conflicts between the militaries of opposing states,"[379] Brooks argues that we should not "romanticize our own species."[380] And even when humans correctly make that distinction, they "often make risk-averse calculations about necessity and proportionality."[381] Finally, the use of VIS may increase the "projection of state power despite declining military recruitment figures."[382] While the use of VIS may result in an arms race for this type of technology, the race should be understood as a *shift* in the means of violence rather than an increase in violence.[383]

In the final analysis, IHL is not exclusively concerned with the "package" of violence—or as one commentator put it, the "who" or the "what"[384]—but it focuses more broadly on the *effects* of violence. IHL is focused on how that violence is "performed" only to the extent it seeks to regulate the effect or impact of a weapon. As Anderson, Reisner, and Waxman note:

> The principle of humanity is fundamental, but it refers, not to some idea that humans must operate weapons, but instead to the promotion of means or methods of warfare that best protect humanity within the

---

[379] *Id.*, at 136.

[380] *Id.*, at 138.

[381] *Id.*, at 136.

[382] Liu, *supra* note 65, at 633.

[383] *See also* DOMINGOS, *supra* note 18, at 281 ("If a war is fought by machines, with humans only in command positions, no one is killed or wounded. Perhaps, then, what we should do, instead of outlawing robots, is – when we're ready – outlaw human soldiers.").

[384] Anderson, Reisner, Waxman, *supra* note 158, at 401 ("Whether the actor on the battlefield is a "who" or a "what" is not truly the issue, but rather how well that actor performs according to the law of armed conflict").

> lawful bounds of war, irrespective of whether the means to that end is human or machine or some combination of the two.[385]

Tensions with International Human Rights Law

Article 6(1) of the International Covenant on Civil and Political Rights—known as the "cornerstone" of International Human Rights Law ("IHRL")—provides that "every human being has the inherent right to life" and that "[n]o one shall be arbitrarily deprived of his life."[386] Whereas IHRL "protects physical integrity and human dignity in all circumstances,"[387] "the law of war allows, or least tolerates, the killing and wounding of innocent human beings not directly participating in an armed conflict, such as civilian victims of lawful collateral damage."[388] The tension between the objectives of IHRL and IHL has led some to argue that "[f]ully autonomous weapons have the potential to contravene the right to life."[389] For example, HRW argues that such weapons will not be able to make the necessary "qualitative assessments" required to comply with IHRL. HRW predicts that "[d]ue to the infinite number of possible scenarios, robots could not be pre-programmed to handle every specific circumstance."[390] It is clear that a machine cannot be pre-programmed in the traditional sense with all eventualities that may arise on the

385 Anderson, Reisner, Waxman, *supra* note 158, at 401.

386 International Covenant on Civil and Political Rights, G.A. Res. 2200A, U.N. GAOR, 21st Sess., Supp. No. 16, at 52, U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171 (entered into force Mar. 23, 1976) [hereinafter ICCPR], art. 6(1).

387 Meron, *supra* note 189, at 240.

388 *Id.*, at 240.

389 Mind the Gap, *supra* note 66, at 16.

390 *Id.*, at 9.

battlefield. But the point ignores the potential of machine learning, as noted above. Machine learning, like a human brain, does not need to be fed all eventualities; rather, humans can teach the machine how to learn and how to make lawful decisions.

Further, from a legal standpoint, IHL, as *lex specialis*, is the governing law in international armed conflicts. As the ICJ noted in *Nuclear Weapons*:

> [W]hether a particular loss of life, through the use of a certain weapon in warfare, is to be considered an arbitrary deprivation of life contrary to Article 6 of the Covenant, can only be decided by reference to the law applicable in armed conflict and not deduced from the terms of the [International Covenant of Civil and Political Rights] itself.[391]

In practical terms, it may be that VIS will be able to utilize non-lethal force to achieve military objectives, and thus still satisfy the spirit IHRL by not using lethal force.

Can VIS Protect the Environment and Cultural Property?

International humanitarian law prohibits attacks on the natural environment[392] and cultural property.[393] VIS may be able to make better determinations about incidental effects on the natural environment than a warfighter because the latter may

---

[391] Nuclear Weapons, *supra* note 230, at 25.

[392] Additional Protocol I, art. 35 (3) ("It is prohibited to employ methods or means of warfare which are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment."); Article 55 ("Care shall be taken in warfare to protect the natural environment against widespread, long-term and severe damage. This protection includes a prohibition of the use of methods or means of warfare which are intended or may be expected to cause such damage to the natural environment and thereby to prejudice the health or survival of the population.").

[393] Additional Protocol, art. 55 ("it is prohibited: (a) to commit any acts of hostility directed against the historic monuments, works of art or places of worship which constitute the cultural or spiritual heritage of peoples;"); *see also* UNESCO Convention for the Protection of Cultural Property in the Event of Armed Conflict, May 14, 1954, 249 U.N.T.S. 240, art. 4(1) ("The High Contracting Parties undertake to respect cultural property situated within their own territory as well as within the territory of other High Contracting Parties by refraining from any use of the property and its immediate surroundings or of the appliances in use for its protection for purposes which are likely to expose it to destruction or damage in the event of armed conflict; and by refraining from any act of hostility, directed against such property.").

be primarily focused on—and overwhelmed by—military objectives. A network of VIS could more quickly process scientific data in order model decisions that minimize long-term and severe damage to the environment. Of course, concepts such as "long-term" or "severe" are hard to define. But environmental science, aided by AI, provides a number of tools designed to predict and quantify damage to the environment.[394] Environmental impact assessments are already used in a number of industries to assess the impact of man-made activities on the natural environment. If faced with an existential threat, would it not be better for a State to use VIS to offset the threat, rather than dropping an atomic bomb that is incapable of distinction and is extremely harmful to the environment? These questions highlight the potential that VIS code can be "green" as well as "good."[395]

VIS could also be used in missions to protect cultural property.[396] VIS may be able to better identify sites protected by international conventions and take action to minimize their damage. Further, since political costs may be too high to send ground forces to protect cultural heritage, sending in VIS may be a more viable alternative.[397]

---

[394] Francois Spitz and Sovan Lek, *Environmental impact prediction using neural network modelling. An example in wildlife damage available*, 36(1) JOURNAL OF APPLIED ECOLOGY 317 (April 1999); Roudgarmi, P., Monavari, M., Feghhi, J. et al., *Environmental impact prediction using remote sensing images*, 9(3) JOURNAL OF ZHEJIANG UNIVERSITY-SCIENCE A, 381-390 (2008).

[395] By "good", I mean code that is capable of powering a weapon system in a way that complies with International Humanitarian Law.

[396] Thiago Velozo and Lucas Bento, *ISIS Is Destroying Priceless Artifacts. Here's How to Stop Them*, THE DIPLOMAT (Mar. 17, 2015) *available at* http://thediplomat.com/2015/03/isis-is-destroying-priceless-artifacts-heres-how-to-stop-them/.

[397] In order to further humanize IHL and align it with IHRL, more research should be focused on technology capable of immobilizing and disarming without lethal consequences.

Human Responsibility in Armed Conflicts

International responsibility under public international law can be broadly divided into two camps: State Responsibility and Individual Responsibility. State Responsibility is a branch of international law governing state responsibility for a State's breaches of its international obligations. Individual Responsibility governs the allocation of personal responsibility under international criminal law. While IHL's capacity to govern VIS has been examined by many,[398] the question of responsibility and accountability for the use VIS is typically ignored or takes a secondary position in the debate.[399] Some argue that no human responsibility ensues.[400] Others propose that responsibility should be shared between robots[401] and the human actors involved.[402] Those perspectives, however, ignore the current legal

[398] *See supra* Chapter III ("Acting Responsibly in Armed Conflicts").

[399] McFarland and McCormack, *supra* note 134, at 385 ("Too little analysis has been undertaken on questions of State responsibility for the deployment of autonomous weapons systems that result in serious violations of the law of armed conflict"); *but see e.g.* McFarland and McCormack, *supra* note 134, at 370 ("[I]n what circumstances, if any, does a system's inbuilt capacity for autonomous operation start to affect the allocation of criminal responsibility for serious violations of the law of armed conflict?").

[400] *See* Mind the Gap, *supra* note 66, at 2 ("These weapons have the potential to commit criminal acts—unlawful acts that would constitute a crime if done with intent—for which no one could be held responsible"); Darren M. Stewart, *New Technology and the Lai of Armed Conflict,* 87 INT'L L. STUD. 271, 290 (2011) ("Absent the aberrant behavior of either the data or command programmers . . . it would be almost impossible to attribute the autonomous system's behavior *directly* to a particular human."); Sparrow, R., *Killer robots,* 41(1) JOURNAL OF APPLIED PHILOSOPHY, 62–77 (2007) (arguing that it will not be possible to hold humans responsible for the behavior of autonomous robots); *id.*, at 62 ("possible loci of responsibility for robot war crimes are canvassed: the persons who designed or programmed the system, the commanding officer who ordered its use, the machine itself. I argue that in fact none of these are ultimately satisfactory.").

[401] Indeed, some have argued that robots in the future be held responsible in some form. Hellstrom, T., *On the moral responsibility of military robots*, 5 ETHICS AND INFORMATION TECHNOLOGY 99 (2013) (arguing that we should be prepared for a future when people blame robots for their actions).

[402] Crnkovic, G. D., & Persson, D., *Sharing moral responsibility with robots: A pragmatic approach*, in P. K. Holst & P. Funk (Eds.), FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS (IOS Press Books, 2008).

architecture of international responsibility. Indeed, humans and human institutions remain solely responsible for VIS-inflicted violence.[403]

State Responsibility

State Responsibility is governed by the International Law Commission's Articles on the Responsibility of States for Internationally Wrongful Acts ("ARISWA").[404] Article 1 of ARISWA provides that international responsibility arises in respect of internationally wrongful acts. Article 2 of ARISWA defines "an internationally wrongful act as occurring when conduct attributable to a state under international law constitutes a breach of an international obligation of the state."[405] Attribution, in turn, is the "process by which international law establishes whether the conduct of a natural person or other such intermediary can be considered an 'act of state', and thus be capable of giving rise to state responsibility."[406] State responsibility rules acknowledge that the State is an abstract entity and can only act through its organs and agents, which includes government personnel and persons

---

[403] Nagenborg, M., Capurro, R., Weber, J., & Pingel, C., *Ethical regulations on robotics in Europe*, AI & SOCIETY (2008), 22, 349–366; Marino, D., & Tamburrini, G., *Learning robots and human responsibility*, 6 INTERNATIONAL REVIEW OF INFORMATION ETHICS, 46–51 (2006); Chopra, S., & White, L. W., A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS, (The University of Michigan Press, 2011).

[404] Articles on the Responsibility of States for Intentionally Wrongful Acts, in Report of the International Law Commission to the General Assembly on Its Fifty-Third Session, 56 U.N. GAOR Supp. No. 10, at 1, 43, U.N. Doc. A/56/10 (2001) [hereinafter ARSIWA].

[405] James Crawford, STATE RESPONSIBILITY: THE GENERAL PART 93 (Cambridge University Press, 2014).

[406] CRAWFORD, *supra* note 405, at 113.

acting under the direction or control of a State.[407] The acts of armed forces, as an organ of the State, are therefore attributable to a State.[408] Thus, if a State's armed forces use VIS in breach of IHL, that State is responsible for the breach.[409]

Acts *ultra vires* are also "attributable to a state where committed by one of its organs,"[410] such as the armed forces.[411] Indeed, "the mere fact a soldier losing his way does not deprive him of the status of a state organ."[412] Thus, a soldier that tampers with a VIS and misuses it does not absolve the State from responsibility.[413]

---

[407] Melzer, *supra* note 5, at 38 ("This includes not only government personnel, such as members of the armed and police forces or intelligence agencies (*de jure* State agents), but also persons acting on the instructions or under the direction or control of a State, such as certain private military or security contractors (*de facto* State agents).").

[408] CRAWFORD, *supra* note 405, at 119 ("The most obvious executive manifestations are actions of the armed forces, which in the context of armed conflict are in all cases attributable to and engage the international responsibility of the state in question.").

[409] *See* ARSIWA, *supra* note 404, art. 12 ("There is a breach of an international obligation by a State when an act of that State is not in conformity with what is required of it by that obligation, regardless of its origin or character."); *Autonomous Weapon Systems under International Law*, ACADEMY BRIEFING NO. 8, Geneva Academy, November 2014 (noting Sweden's position that "even if LAWS are referred to as "autonomous", states are legally responsible for their use. If violations occur that are attributable to a particular State, that state is responsible according to the rules of State responsibility and international humanitarian law.'").

[410] CRAWFORD, *supra* note 405, at 120; ARSIWA, *supra* note 404, art. 7 ("The conduct of an organ of a State or of a person or entity empowered to exercise elements of the governmental authority shall be considered an act of the State under international law if the organ, person or entity acts in that capacity, even if it exceeds its authority or contravenes instructions.").

[411] If internal laws grant robots and VIS some type of legal status, so that it becomes some form of "entity", the State would remain responsible for the acts of that entity. *See* ARSIWA, *supra* note 404, art. 5 ("The conduct of a person or entity which is not an organ of the State under article 4 but which is empowered by the law of that State to exercise elements of the governmental authority shall be considered an act of the State under international law, provided the person or entity is acting in that capacity in the particular instance.").

[412] CRAWFORD, *supra* note 405, at 120.

[413] Melzer, *supra* note 5, at 38 ("In international armed conflict, the responsibility of States even extends to 'all acts committed by persons forming part of its armed forces,' including acts committed outside their official capacity as members of the armed forces. In principle, therefore, all military operations carried out on behalf of a State are directly attributable to that State, regardless of where they take place or where their effects materialize.").

Depending on the misuse, individuals would also be responsible for that conduct. Further, "a state cannot plead the provisions of its internal law in order to justify an internationally unlawful act."[414]

International legal responsibility may also arise where a State aids or assists another State to commit an internationally wrongful act, such as supplying the code or hardware to the assisted State. While the assisting State need not *know* of the unlawfulness of the assisted conduct, it must be aware of the "factual circumstances" that make it unlawful.[415] For example, when a State is known to use VIS in contravention to IHL, other States knowingly assisting such operations through the provision of code, training, or hardware will be internationally responsible for their assistance.[416]

Can VIS' autonomy breach the chain of responsibility? In the *Nicaragua* case, in considering Guatemala's argument that the *contras* were mercenaries paid by the US, the ICJ noted that the *contras* would "have no real autonomy in relation to [US]" and as such "any offences which they *have committed* would be imputable to [the US]."[417] This passage of obiter dicta may at first imply that autonomy can undermine the chain of responsibility. However, closer analysis shows how the ICJ was referring to a group's relationship to a State, not a weapon's status in relation to

---

[414] CRAWFORD, *supra* note 405, at 121.

[415] Melzer, *supra* note 5, at 38.

[416] *Id.* ("For instance, when a State is known to carry out armed drone attacks which are widely regarded to contravene its obligations under human rights law and humanitarian law, other States knowingly assisting such operations through the provision of personnel, logistic support or targeting intelligence will become internationally responsible for doing so.").

[417] Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States), ICJ Reports (1986) 14, 64 (emphasis added).

the State. Further, a weapon cannot "commit an offence" under international law—only States and individuals i.e., humans and human institutions, are capable of violating international law.

One of the most significant challenges to State Responsibility in the context of VIS is the defense of *force majeure*. Article 23(1) of ARISWA states:

> The wrongfulness of an act of a State not in conformity with an international obligation of that State is precluded if the act is due to force majeure, that is the occurrence of an irresistible force or of an unforeseen event, beyond the control of the State, making it materially impossible in the circumstances to perform the obligation.

The concern is that "[w]ith increasing autonomy" weapon systems will "spin[] out of control" and "malfunction."[418] A State could therefore in theory plead *force majeure* in order to "evade" international responsibility for "unforeseen" VIS breaches if IHL, such as indiscriminate attacks on civilians. However, ARISWA, Article 32(2)(a), provides that force majeure is not available where "the situation of force majeure is due, either alone or in combination with other factors, to the conduct of the State invoking it." As the Commentary to the ARISWA notes, "material impossibility cannot be invoked if the impossibility is *the result of a breach by that party* either of an obligation under the treaty or of any other international obligation owed to any other party to the treaty."[419] Thus, if a State failed to take the necessary precautions to ensure that the VIS complied with IHL (lack of training commanders and officers, lack of testing, poor development), then that State would not be able to plead *force majeure* if the VIS malfunctioned or did not operate as planned. Indeed, States

[418] Melzer, *supra* note 5, at 38.

[419] ARSIWA, *supra* note 404, art. 23, § 9.

cannot invoke *force majeure* by relying on events resulting from the state's negligence.[420] In any event, the effects of man-made code or robotic systems are not "irresistible forces" and potential malfunctioning are not "unforeseen events."

Further, Article 23(2)(b) of ARISWA provides that a State cannot plead *force majeure* if "the State has assumed the risk of that situation occurring." In this context, Crawford explains that "[i]f a state accepts responsibility for a particular risk, it renounces its right to rely on *force majeure* to evade that responsibility. It may do so expressly, by agreement, or by clear implication."[421] When a State assumes the risk of deploying a VIS that runs ineffective or faulty software or hardware, that State is responsible for the unexpected operation of that VIS because "there is predictable unpredictability" in the use of VIS.[422] Thus, a State that employs these technologies in warfare and seek to benefit from them takes on the risks associated with their use.[423] A thornier question is whether external hacking of the VIS constitutes *force majeure*.[424] But hacking is a well-known countermeasure in cyberspace, and thus a

---

[420] *See* CRAWFORD, *supra* note 405, at 298 ("One thing it certainly does not cover is mere negligence.").

[421] CRAWFORD, *supra* note 405, at 301.

[422] Robin Geiss, *Autonomous Weapons Systems: Risk Management and State Responsibility*, Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 2016) *available at* http://www.unog.ch/80256EDD006B8954/(httpAssets)/00C95F16D6FC38E4C1257F9D0039B84D/$file/Geiss-CCW-Website.pdf.

[423] Geiss, *supra* note 422, at 2 ("It follows that a State that benefits from the various (strategic) gains associated with this new technology should therefore be held responsible whenever the (unpredictable) risks inherent in this technology are realized.").

[424] Melzer, *supra* note 5, at 39 ("The question to be clarified is whether *force majeure* can only arise if robotic malfunction is caused by external factors (e.g., malicious cyber-interference) or also in case of a genuinely unforeseen shortcomings in the system's operating software.").

risk that a State assumes in entering the cyberspace domain. Thus, even in the most extreme circumstances, a State remains responsible for the use of VIS.[425]

Individual Responsibility

Individual responsibility is governed by international criminal law ("ICL"). ICL revolves around the notion of personal culpability,[426] which can be direct or complicit.[427] Personal culpability, in turn, is informed by individual autonomy. As Cassese notes:

> The principle of individual autonomy whereby the individual is normally endowed with free will and the independent capacity to choose his conduct is firmly rooted in modern criminal law, including ICL.[428]

Human autonomy, however, is not to be confused with the notion of VIS' autonomy. The latter is merely a product of human choice.[429] The autonomous operation of a weapon system does nothing to undermine a human being's autonomy to develop,

---

425 Melzer, *supra* note 5, at 37 ("At the most basic level, legal accountability requires the recognition that States remain legally responsible for the consequences of their use of robotic weapons irrespective of the operational autonomy achieved by such systems. Legal accountability also involves a governmental duty of investigation, and of reparation for potential violations.").

426 CASSESE, *supra* note 204, at 33 ("In ICL, the general principle applies that no one may be held accountable for an act he has not performed or in the commission of which he has not in some way participated, or for an omission that cannot be attributed to him.").

427 DoD Law of War Manual, *supra* note 197, ¶ 18.32 ("Individuals may be held liable for violations of the law of war whether they have committed them directly or are complicit in the commission of such crimes.").

428 CASSESE, *supra* note 204, at 33.

429 Schmitt and Thurnher, *supra* note 54, at 280 ("Finally, humans will always be accountable for the employment of autonomous weapon systems. Although they will gradually delegate more tasks to autonomous systems, the responsibility for the appropriate use of the systems will nevertheless remain with the human operators and commanders. Orders to deploy the system and judgments about how to program it will come from a human. Any recklessness or criminal misuse will result in accountability through the same war crimes mechanisms that already exist under the law of armed conflict.").

select and deploy weapons to carry out human objectives. What better evidence of human free will than to remove a human being from harm's way and let a machine do the work on its behalf?

Individual criminal responsibility can be divided in two camps: direct responsibility and indirect responsibility.[430] Individuals that develop and deploy VIS with the intention to commit international crimes will be *directly* responsible for those crimes. For example, under the doctrine of direct responsibility, military commanders who order their subordinates to perpetrate atrocities in violation of the law are liable for the ensuing crimes.[431] Thus, if a commander ordered subordinates to deploy VIS programmed to commit war crimes, then that commander, as well as the subordinates, would be liable for those crimes. Accordingly, "[s]uperiors shall only issue orders which are in conformity with international law. Superiors who issue an order contrary to international law expose not only themselves but also their subordinates obeying these orders to the risk of being prosecuted."[432]

---

[430] Wagner, *supra* note 39, at 35 ("Given that AWS are military tools, a natural starting point for responsibility could be the military officers who set parameters for a given engagement. It is important to distinguish between direct responsibility, which arises from acts or omissions supporting the commission of IHL, and command responsibility, which involves the failure of military or civilian superiors to conduct the required oversight of their subordinates.").

[431] *See e.g.* Rome Statute of the International Criminal Court art. 25(3)(b), July 17, 1998, 2187 U.N.T.S. 90 (recognizing this mode of criminal liability); O'Connell, *supra* note 190, at 39 ("In the case of *Yamashita*, the US Supreme Court held that General Yamashita was guilty of a war crime for failing to control the troops under his command and to prevent the atrocities which they committed in areas occupied by the Japanese army. This principle has now been incorporated into the leading texts on international criminal law such as the Statute of the International Criminal Court (Articles 25 and 28) and the International Criminal Tribunal for the Former Yugoslavia (Article 7).").

[432] O'Connell, *supra* note 190, at 40 ("An officer, of whatever rank, who orders the commission of an unlawful act is guilty of a war crime, as is the soldier who carries out that order. The 'grave breaches' provisions of the Geneva Conventions and AP I stipulate that ordering the commission of an act amounting to a grave breach is itself a grave breach."); Additional Protocol I, art. 86; *Entrenching Impunity Government Responsibility for International Crimes in Darfur*, HUMAN RIGHTS WATCH, *available at* https://www.hrw.org/sites/default/files/features/darfur/fiveyearson/report9.html ("All individuals, including government officials, military commanders,

Commanders may also be liable under the doctrine of indirect responsibility, which states that a commander is liable when the commander knew or should have known[433] that her *subordinates* committed crimes and the commander failed prevent,[434] report, or punish such subordinates.[435] Critics argue that this doctrine cannot be squarely applied to the use of VIS. Their arguments tend to be based on a false comparison that equates a VIS to a "subordinate" or "combatant."[436] This

---

soldiers, militia members, and civilians, are subject to prosecution for war crimes, crimes against humanity, and applicable domestic crimes under international law.").

433 *Command responsibilty and failure to act*, Advisory Service on International Humanitarian law INTERNATIONAL COMMITTEE OF THE RED CROSS, 1, *available at* https://www.icrc.org/eng/assets/files/2014/command-responsibility-icrc-eng.pdf.

434 The "abort" switch mentioned above may provide a mechanism to ensure that commanders can prevent any unlawful violence inflicted by a VIS. However, it is acknowledged that some VIS may act at speeds that may complicate real-time human monitoring. This only becomes a problem, however, if the technology is incompetent from the beginning. It is unclear how or why a military will deploy a weapon they do not understand or do not have confidence in. For both legal and strategic reasons, such weapons are unlikely to be deployed at all.

435 Additional Protocol I, art. 87 (1) ("Parties to the conflict shall require military commanders, with respect to members of the armed forces under their command and other persons under their control, to prevent and, where necessary, to suppress and to report to competent authorities breaches of the Conventions and of this Protocol.").

436 *See e.g.* Mind the Gap, *supra* note 66, at 20 ("Using a commander-subordinate analogy, the commander would not be directly responsible for the robot's specific actions since he or she did not order them."); *id.*, at 21 ("First, command responsibility only arises when a subordinate commits a chargeable criminal offense. Second, even if a criminal act committed by a robot were considered sufficient for the command responsibility doctrine in the case of fully autonomous weapons"); Heyns, *supra* note 7, at 15 ("Since a commander can be held accountable for an autonomous human subordinate, holding a commander accountable for an autonomous robot subordinate may appear analogous."); Roff *supra* note 86, at 357 ("Military commander liable for actions of subordinates, where commander should have known what would happen. Modern case law requires the effective control of subordinates by superiors . . . . Looking at the case of LARs, the effective control criterion would actually exculpate leaders from legal responsibility because the commanders' inability to control the machines. Autonomous machines are "impossible" to control by "a human in real-time due to its processing speed and the multitude of operational variables involved."); Neha Jain, *Autonomous weapons systems: new frameworks for individual responsibility*, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, at 303 ("The actions of an AWS, being partly of the character of a weapon and partly the character of the combatant"); Liu, *supra* note 65, at 636 ("the capacity for autonomous decision-making pushes these technologically advanced systems to the boundary of the notion of 'combatant'").

misleading characterization of the weapon system as some type of robotic soldier[437] or subordinated combatant[438] ultimately tips the scales of criticism against the efficacy of international law. To be clear, VIS are weapon systems—not "subordinates" or "combatants." A weapon is not a subordinate and a weapon cannot commit an international crime as that term is understood under international law. Indeed, in the words of the Nuremberg judges, "crimes . . . are committed by men, not by abstract entities."[439]

Further, critics argue that commanders may not be held responsible because they would not understand how the VIS' complex programming operates and thus could not have "known" something they did not comprehend.[440] But the commander need not be an expert in computer science or know the intricacies of how the system is built, only what it is "able and unable to do."[441] To this end, DARPA is working

---

[437] Ugo Pagallo, *Robots of just war: a legal perspective*, 24 PHILOSOPHY AND TECHNOLOGY, 307– 323 (2011).

[438] *See e.g.* KRISHNAN, *supra* note 255, at 103 ("AW could potentially interrupt the clear chain of military command that is required by international law. Article 1 of the Hague Convention requires *any combatant* 'to be commanded by a person responsible for his *subordinates*'") (emphasis added); Liu, *supra* note 65, at 636 ("the capacity for autonomous decision-making pushes these technologically advanced systems to the boundary of the notion of 'combatant'.").

[439] 1 TRIAL OF THE MAJOR WAR CRIMINALS BEFORE THE INTERNATIONAL MILITARY TRIBUNAL 223 (1947) ("Crimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced.").

[440] Mind the Gap, *supra* note 66, at 3 ("a commander would not always have sufficient reason or technological knowledge to anticipate the robot would commit a specific unlawful act."); Heyns, *supra* note 7, at 15 ("Yet traditional command responsibility is only implicated when the commander "knew or should have known that the individual planned to commit a crime yet he or she failed to take action to prevent it or did not punish the perpetrator after the fact." It will be important to establish, inter alia, whether military commanders will be in a position to understand the complex programming of LARs sufficiently well to warrant criminal liability.").

[441] Sassòli, *supra* note 150, at 324 (added emphasis); Schmitt and Thurnher, *supra* note 54, at 267 ("This requires an understanding not only of the physical capabilities and limitations of the system (the maximum range, the effectiveness of the weaponry, the blast radius of its weapons, etc.), but also the subjective values embedded in it"); Schmitt and Thurnher, *supra* note 54, at 277 ("The mere fact

on ways to make "the decisions made by autonomous systems to use lethal weapons 'explainable'" so that humans can have more details about "how the machine used deep learning to come up with answer[s] [to problems]."[442]

The more challenging question is whether individuals can be held responsible for unlawful violence inflicted by VIS that are "unintended" or "unknowable," either because of malfunctioning or external hacking. These individuals may argue that such violence was unintended and unforeseeable, and that they took all precautions to ensure that only lawful violence was used. Critics argue that such individuals may argue that they lacked the necessary *mens rea* to commit international crimes[443] and thus evade liability. But this concern is overestimated for five reasons.

First, there is little doubt that collective responsibility i.e., State responsibility would arise in these circumstances.[444] Article 91 of the Additional Protocol I provides that a State "shall be responsible for all acts committed by persons forming part of its armed forces." Further, Articles 86 and 87 of Additional Protocol I Additional Protocol I explicitly impose an international legal obligation on parties to

---

that a human might not be in control of a particular engagement does not mean that no human is responsible for the actions of the autonomous weapon system. A human must decide how to program the system and when to launch it. Self-evidently, that individual would be accountable for programming it to engage in actions that amounted to war crimes. Moreover, the commander or civilian supervisor of the person would be accountable for those war crimes if he or she knew or should have known that the autonomous weapon system had been so programmed and did nothing to stop its use, or later became aware that the system had been employed in a manner constituting a war crime and did nothing to hold the individuals concerned accountable.").

[442] Cameron Leuthy, *In battle, you need to trust your robots*, BLOOMBERG GOVERNMENT (Jun. 16, 2016) *available at* http://about.bgov.com/blog/battle-need-trust-robots/.

[443] *See e.g.* Mind the Gap, *supra* note 66, at 21("Command responsibility is only triggered if a commander has actual or constructive knowledge of the crime, that is, the commander must know or have a reason to know of the crime.").

[444] *See* Additional Protocol I, art. 91 ("A Party to the conflict which violates the provisions of the Conventions or of this Protocol shall, if the case demands, be liable to pay compensation. It shall be responsible for all acts committed by persons forming part of its armed forces.").

an armed conflict to require commanders to take all necessary measures[445] to prevent crimes being committed and to initiate disciplinary or penal sanctions against them when crimes have been perpetrated.[446]

Second, military leaders and commanders' responsibilities constrain[447] the potential unpredictabilities, if any, of using Violent Intelligent Systems. Indeed, before engaging an objective, military leaders must comply with a number of international obligations, including selecting means and methods of warfare to minimize collateral injury to civilians,[448] refraining from launching excessive force which may be expected to cause incidental injury civilians,[449] and suspending attacks if new information suggests that the target is not a military objective.[450] These responsibilities add layers of checks and balances over possible uncertainties related to the use of VIS. Further, in addition to "sign[ing] off" on decisions, commanders and military leaders are responsible for "formulating the rules of engagement, which includes specifying how and which weapons may be used."[451] Thus commanders may be reluctant to allow subordinates to use a VIS that "they know is unpredictable

[445] Additional Protocol I, art. 86 (2) ("The fact that a breach of the Conventions or of this Protocol was committed by a subordinate does not absolve his superiors from penal or disciplinary responsibility, as the case may be, if they knew, or had information which should have enabled them to conclude in the circumstances at the time, that he was committing or was going to commit such a breach and if they did not take all feasible measures within their power to prevent or repress the breach.").

[446] Additional Protocol I, arts. 86, 87.

[447] Noorman and Johnson, *supra* note 122, at 59.

[448] Additional Protocol I, art. 57 (2)(a)(ii).

[449] Additional Protocol I, art. 57 (2)(a)(iii).

[450] Additional Protocol I, art. 57 (2)(b); Fleck, *supra* note 188, at 199.

[451] Noorman and Johnson, *supra* note 122, at 59.

for fear that they would be held responsible for violating the laws of armed conflict as a result of the robot's rogue or unethical behavior."[452]

To this end, military policy may emphasize direct individual responsibility and accountability for the operation of VIS.[453] For example, US military policy imposes direct obligations on persons responsible for making decisions related to the use of autonomous weapon systems.[454] Indeed, Directive 3000.09 provides that "[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement."[455] This creates a direct thread of human responsibility and judgment through the employment of VIS capable of autonomous operation.

---

[452] *Id*.

[453] *See e.g.* UK Joint Doctrine, *supra* note 257, at ¶ 510 (as long as a system is lawful, "[l]egal responsibility for any military activity remains with the last person to issue the command authorising a specific activity."); *but see* Liu, *supra* note 65, at 650 ("Relocating the locus of punishment to natural persons with the closest nexus to these machines, however, runs the risk of scapegoating those persons: the possession of autonomous decisionmaking capacity may break the causal chain that would justify the attribution of responsibility to those persons. Thus, autonomous and remote weapons systems may have a higher capacity to adhere to IHL, but will inevitably have much lower levels of responsibility for any breaches. This leads to impunity for conduct in armed conflict.").

[454] Directive Number 3000.09, *supra* note 35, at ¶ 4.(b) ("[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement."); Directive Number 3000.09, *supra* note 35, Enclosure 4, sec. 10 (imposing these same requirements on the "Commanders of the U.S. Combatant Commands.").

[455] Directive Number 3000.09, *supra* note 35, at ¶ 4.(b) ("[p]ersons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement."); *id.*, at Enclosure 4, sec. 10 (imposing these same requirements on the "Commanders of the U.S. Combatant Commands.").

Third, the possibility that human operators *may* lack the necessary *mens rea* for war crimes in this context should not act as a basis for prohibiting these types of weapons.[456] Any weapon can malfunction. Why should we assume that the human operator acted benevolently and lawfully? Why are we assuming that the VIS malfunctioned, and not the other way around? While the discourse must recognize potential pitfalls of the technology, it should not fuel narratives that allow humans to hide behind the complexity of their technologies.

Further, humans can lie about their intentions, but code seldom lies. Human intention can be evidenced in many of the components and settings of a VIS, including in code that operates it and the type of weapons it employs. If one uses VIS to commit crimes against humanity, the VIS must, for example, be programmed to commit murder against civilians. Of course, sophisticated coding techniques, such as encryption and data self-destruction technology, may conceal evidence of premeditation or knowledge for *mens rea* purposes. In cases where the VIS may have genuinely malfunctioned, code may nonetheless illustrate human failures in setting appropriate parameters for the mission at hand, by for example setting lower thresholds of doubt in environments heavily populated with civilians. Here, experts would need to opine on whether specific code and models were set to target legitimate targets in lawful ways. In any event, VIS code, as a symbolic language of

---

[456] *See e.g.* Mind the Gap, *supra* note 66, at 2 ("Human commanders or operators could not be assigned direct responsibility for the wrongful actions of a fully autonomous weapon, except in rare circumstances when those people could be shown to have possessed the specific intention and capability to commit criminal acts through the misuse of fully autonomous weapons. In most cases, it would also be unreasonable to impose criminal punishment on the programmer or manufacturer, who might not specifically intend, or even foresee, the robot's commission of wrongful acts.").

violence, can evince human intent to inflict violence. Ultimately, whether that code (or the VIS hardware) malfunctioned is a question of fact.

Fourth, technological developments will facilitate the enforcement of responsibility. This thesis has already considered how built-in monitoring systems will help assess VIS compliance with IHL. This could be taken a step further to include a traceable path of command within the system, so that "every distinct decision can be traced back to the responsible individual, who may then be held accountable."[457] Enemy surveillance of VIS operations will also add a layer of accountability. Indeed, the enemy may publish footage of VIS violations captured from its surveillance apparatus.

Finally, the question of (and doubts concerning) human responsibility will become less of an issue once there is a better understanding of how VIS, and AI generally, will operate. To this end, additional interdisciplinary efforts must be pursued to facilitate *cross-disciplinary* research into VIS to reach a consensus about the potential of their underlying technologies as well as issues related to terminology, legality, and best practices. The real issues are practical: how can we ensure that VIS are legally and technically competent? How can we adequately test them? Can we develop international best practices for their development, testing, and fielding in order to minimize risks? How can armed forces share technologies (e.g., open-source code) without undermining competitive advantage?[458] How can we properly train

---

[457] Wagner, *supra* note 39, at 38.

[458] *See* Russell, Dewey, and Tegmark, *supra* note 314, at 107 ("If it is permissible or legal to use lethal autonomous weapons, how should these weapons be integrated into the existing command-and-control structure so that responsibility and liability be distributed, what technical realities and forecasts should inform these questions, and how should "meaningful human control" over weapons be defined?").

human actors involved in the development and deployment of VIS?[459] How can the international community monitor the use of VIS?[460] Ultimately, the legal evaluation of the employment of any weapon, including who is to be responsible for its employment, will depend on its specific use.[461]

[459] Noorman and Johnson, *supra* note 122, at 60 ("when the black box is opened up and we see how autonomy is understood and 'made' by those involved in the design and development of robots, the responsibility questions change significantly. The important question is not whether human actors can be held responsible (they can), but how tasks are distributed among human and non-human components of the system, whether the machine parts have been adequately tested, whether the human actors involved have been adequately trained for their tasks, what risks are involved, and how those risks are being managed and minimized.").

[460] Lewis, Blum, and Modirzadeh, *supra* note 6, at ix (noting the role of scrutiny governance, which "concerns the extent to which a person or entity is and should be subject to, or should exercise, forms of internal or external scrutiny, monitoring, or regulation (or a combination thereof) concerning the design, development, or use of a war algorithm. Scrutiny governance does not hinge on—but might implicate—potential and subsequent liability or responsibility (or both). Forms of scrutiny governance include independent monitoring, norm (such as legal) development, adopting nonbinding resolutions and codes of conduct, normative design of technical architectures, and community self-regulation").

[461] *See* Fleck, *supra* note 188, at 182 ("As stated by Special Rapporteur Philip Alston in his report on extrajudicial, summary, or arbitrary executions in 2010, 'a missile fired from a drone is no different from any other commonly used weapon, including a gun fired by a soldier or a helicopter or gunship that fires missiles'. He stressed that the legal evaluation of the employment of any weapon depends on its specific use. Outside an armed conflict, the use of lethal force must comply with human rights limitations. The fact that UAVs/UCAVs have no on-board pilot does not change the applicability of the principles of armed conflict to UAVs/UCAVs.").

Chapter IV

Networks of Human Violence

Re-Imagining Human Violence in International Armed Conflicts

The hostility towards the use of VIS in international armed conflicts is fundamentally grounded on a perception that autonomous technology will allow violence to be unleashed uncontrollably and unlawfully. Taken to its limit, the critics' interpretation suggests that international law is only capable of regulating *human* means and methods of violence. In other words, the use of other intelligent actants in the infliction of violence could fall outside of international law's reach, despite being designed, developed, directed and deployed by humans to achieve human objectives. In addition to misinterpreting the law, this perception misconceptualizes the structure of human violence in international armed conflicts.

Human violence is not exclusively performed by humans. Rather, it is performed by a *network* of human and non-human actants[462] where responsibilities are distributed—not delegated[463]—to achieve political goals. This networked approach to violence has become increasingly relevant at a time where technology is

[462] Walter Isaacson called the collaboration between machines and humans the "Lovelace approach," named after Ada Lovelace, the English mathematician and crucial work on one of the first computers, the Analytical Engine. Isaacson notes that the Lovelace approach involves machines never "truly think[ing]" and that "humans will always provide creativity and intentionality." Walter Isaacson, *The Intersection of the Humanities and the Sciences*, Lecture, National Endowment for the Humanities (2014) *available at* https://www.neh.gov/about/awards/jefferson-lecture/walter-isaacson-lecture.

[463] *See* Sharkey, *supra* note 260, at 23 ("There is an ongoing technological transformation in warfare with ever more control of weapons being delegated to computer systems.").

becoming deeply embedded in most facets of human activity. The U.S. military has recognized as much, and the Pentagon is preparing for what it is now calling "network-on-network warfare" against traditional rivals, including China and Russia.[464] These networks are founded on "human-machine collaboration"[465] that seeks to "put learning machines, AI, and autonomous systems into the network to allow your network to prevail over an enemy's network."[466] To this end, DARPA is conducting research to enhance collaboration between human beings and machine.[467] For example, swarming technology will enable the use of multiple VIS as "force multipliers" without requiring constant human supervision. As more militaries continue to invest in robotics and artificial intelligence, the performance of human violence will be increasingly distributed within networks of human and non-human actants.

---

[464] Davenport, *supra* note 46 ("The Pentagon is preparing for what Deputy Defense Secretary Robert Work called "network-on-network warfare" against more traditional rivals, such as China and Russia, after more than a decade of counterinsurgency warfare in Iraq and Afghanistan.").

[465] Davenport, *supra* note 46; *see also* Heyns, *supra* note 7, at 9 ("Their most likely use during armed conflict is in some form of collaboration with humans, although they would still be autonomous in their own functions.").

[466] Bryant Jordan, *Pentagon Wants Artificial Intelligence to Defeat Enemy Networks*, DEFENSE TECH (May 5, 2016) *available at* http://www.defensetech.org/2016/05/05/pentagon-wants-artificial-intelligence-to-defeat-enemy-networks/.

[467] For example, the DoD is researching how "to develop autonomous swarming algorithms for Unmanned Vehicle (UxVs) to augment ground troops performing missions in a complex environment, without creating a significant cognitive burden, . . . [and using] minimum operator training and supervision so that the operator can continue to perform his/her normal duties while using UxVs as force multipliers." DARPA 2016 Budget, *supra* note 48, at 134. The Mobile Infantry Program seeks to develop "a system-based, mixed team of mounted/ dismounted warfighters and semi-autonomous variants of current or planned small off-road platforms (equivalent to high-mobility platforms currently used by special forces operators single rider, two-rider, or four-rider variants)." DARPA 2016 Budget, *supra* note 48, at 127 ("The MI mixed teams will be able to execute an expanded mission set from those currently employed. The MI system concept will allow for a combined set of mounted and dismounted operations and for a larger area of operations over more aggressive timelines than standard infantry units.").

But the distribution of practical responsibilities between man and machine does not result in the distribution of legal responsibilities. The assumption that legal responsibilities travel with practical responsibilities runs counter to the networked structure of human violence.[468] Networks of human violence are first and foremost man-made networks.[469] While some theories have sought to portray humans as "no more than pieces of a larger military-industrial machine,"[470] others more precisely recognize that human beings are at the heart of violent networks. The technologies humans employ in these networks are not "self-acting" but are rather operating *in collaboration* with people. Indeed, as Mindell notes, "automated and autonomous vehicles constantly return[] to their human makers for information, energy, and guidance."[471] Ultimately, VIS are "embodi[ments] of human efforts"[472] and products of human intentionality. The characterization of machines, robots, or intelligent

[468] The law is replete with devices that illustrate the notion that legal responsibilities need not travel with practical responsibilities. For example, state responsibility recognizes that, while human agents may hold practical responsibilities for certain acts, states remain responsible for those acts. Command responsibility similarly provides for hierarchical accountability. The doctrine of *respondeat superior* acknowledges that companies, while distributing practical responsibilities through human agents, only attribute certain legal responsibilities to specific human principals. Products liability law generally provides that makers of products are responsible for injuries causes by the products. The common law maxim *qui facit per alium facit per se* provides that "he who acts through another acts himself operates to make the acts of an agent." *See e.g.* Colonial Sec., Inc. v. Merrill Lynch, Pierce, Fenner & Smith Inc., 461 F. Supp. 1159, 1165 (S.D.N.Y. 1978)("the common law maxim Qui facit per alium facit per se, he who acts through another acts himself operates to make the acts of an agent within the scope of his authority, in legal effect, the acts of his principal.").

[469] MINDELL, *supra* note 126, at 10 ("How a system is designed, by whom, and for what purpose shapes its abilities and its relationships with the people who use it.").

[470] DELANDA, *supra* note 95, at 3.

[471] MINDELL, *supra* note 126, at 4; *but see* Liu, *supra* note 65, at 650. Citing M Coeckelbergh at 273 ("None of the parts, nodes, or bees control the action (in this sense they are not agents), but the system, network, or swarm as a whole acts.").

[472] MINDELL, *supra* note 126, at 220.

systems as "inhuman" is thus also incorrect because they are "embedded, [as humans are], in social and technical networks" created by humans.[473]

The notion of machines replacing humans on the battlefield is a "myth" because "[f]or any apparently autonomous system, we can always find the wrapper of human control that makes it useful and returns meaningful data."[474] Technology does not replace humans in the sense that it substitutes them; rather it re-places in the sense that it resituates or repositions the human within the network.[475] In its report entitled *The Role of Autonomy in DoD Systems*, the DoD notes:

> Treating autonomy as a widget or "black box" supports an "us versus the computer" attitude among commanders rather than the more appropriate understanding that there are no fully autonomous systems just as there are no fully autonomous soldiers, sailors, airmen or Marines. Perhaps the *most important message for commanders is that all systems are supervised by humans to some degree*, and the best capabilities result from the coordination and collaboration of humans and machines.[476]

The key to remedying the perspective that humans are "out of the loop" is to focus on the system. As Mindell notes, "[t]here is a human being in the system. *The human being is what makes the system*."[477] Indeed, humans "are still present inside" the VIS

---

[473] *Id.*, at 8.

[474] *Id.,* at 9.

[475] *Id.*, at 13 (Where are the people? Which people are they? What are they doing? When are they doing it?).

[476] *The Role of Autonomy in DoD Systems*, Task Force Report, UNITED STATES DEPARTMENT OF DEFENSE, DEFENSE SCIENCE BOARD (Jul. 2012) *available at* http://fas.org/irp/agency/dod/dsb/autonomy.pdf at 24 (emphasis added).

[477] MINDELL, *supra* note 126, at 220 (emphasis added).

"through design and coding."[478] In effect, code enables "human action" and human intention "removed in time."[479]

The idea that "human decisions, presence, and expertise . . . *still*"[480] lives through another medium does not require any futuristic thought experiments. We can find examples right here on Earth. In *The Extended Phenotype*, evolutionary biologist Richard Dawkins makes the case that nature is full examples of an animal's genes influencing the development of functional extensions of that animal's body. As Dawkins put it, "[a]n animal's behaviour tends to maximize the survival of the genes 'for' that behaviour, whether or not those genes happen to be in the body of the particular animal performing it."[481] For example, a spider's web and a beaver dam are phenotypic extensions of the spider's and the beaver's genes, respectively. "In principle," Dawkins argues, "there is no reason why the phenotypic levers of gene power should not reach out for miles. A beaver dam is built close to the lodge, but the effect of the dam may be to flood an area thousands of square metres in extent."[482]

---

[478] *Id.*, at 10.

[479] *Id.*, at 221 ("This is the essence of the term 'programming' – telling a computer what to do at some point in the future, when the program is run. Of course the machine will respond to its environment, and may encounter novel situations, and may even develop unexpected behaviors. But the constraints on those behaviors are still very tight, and very much pre-scripted by the designers and programmers."); *but* see Heyns, *supra* note 242, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, 4 ("The increased autonomy in weapons release now points to an era where humans will be able to be not only physically absent from the battlefield but also psychologically absent, in the sense that computers will determine when and against whom force is released.").

[480] *Id.*, at 224 ("[H]uman decisions, presence, and expertise are still there but shifting with new technologies, although not always in the ways we expect. It is not the robots themselves, but the novel mixtures of human and automated machines that are changing the nature of the work and the people who do it.") (emphasis added).

[481] RICHARD DAWKINS, THE EXTENDED PHENOTYPE 233 (Oxford University Press, 1999).

[482] *Id.*, at 200.

Similarly, technology is the phenotypic expression of human beings.[483] VIS, then, are phenotypic expressions of human violence. That the technology is removed in space and time from the human does not undermine its human objectives just as the predatory effect of a spider is not undermined simply because the spider web exists outside the spider's body.[484] VIS and humans are linked to "vast networks of data, colleagues, and imagery even as they become enmeshed in the details of human events that unfolds halfway across the globe."[485] Code is thus an extension of the human military organism.[486] To this end, the intelligence present in Violent Intelligent Systems is a functional extension of human predatory will.[487] The desire to build VIS could also be interpreted as motivated by self-preservation. Indeed, by sending VIS to do some of the fighting on their behalf, human lives are spared.[488]

Another way to look at networks of human violence is through the lenses of active externalism. With the advent of technology, human beings have heavily relied

---

[483] DOMINGOS, *supra* note 18, at 284; MINDELL, *supra* note 126, at 13 (technology allows "[p]eople's minds [to] tavel to other places, other countries, other planets.").

[484] DAWKINS, *supra* note 481, at 200 ("In a very real sense her web is a temporary functional extension of her body, a huge extension of the effective catchment area of her predatory organs.").

[485] MINDELL, *supra* note 126, at 220.

[486] To argue that autonomous systems are completely autonomous and do not have any "programmed mission" undermines the bedrock of military action. Soldiers are not given *carte blanche* to do as they please on the battlefield, or to pursue their own private or divergent interests, but military missions are always governed by specific commands, orders and instructions. Similarly, VIS will be given specific instructions on the battlefield. Code, as an extension of human intent, will govern the VIS' actions.

[487] DOMINGOS, *supra* note 18, at 284 ("intelligence and will may not inhabit the same body, provided there is a line of control between them").

[488] I owe this analogy to Thiago Bento's observation that the development of artificial intelligence is an extension of human intelligence and may be a way for human genes' to promote self-preservation. Indeed, if we are able to upload our consciousness on a machine, Man potentially becomes immortal. This may the ultimate expression of self-preservation in a known species.

on environmental supports to perform tasks which otherwise would have been impossible or inefficient for humans to do. Humans use the "general paraphernalia of language, books, diagrams, and culture" to allow their brains to perform some operations, "while others are delegated to manipulations of external media."[489] For example, humans use pens and paper (or calculators) to perform mental calculations. In this sense, technology acts as a cognitive aid to human beings. VIS are thus external media used by humans to carry out violent intentions in international armed conflicts. As gatherers of data and information processors, VIS are cognitive aids that enable human intentions to be implemented remotely with greater efficiency, precision, and effectiveness. Together with humans, VIS form a "coupled system" where all components play "an active causal role", and as such "are *in the loop*, [and] not dangling at the other end of a long causal chain."[490] Eventually, the use of VIS, as a cognitive tool of violence, will influence how humans wage and organize wars. Like pens and calculating machines impacted information technology, VIS will facilitate the human processing of power. In particular, code, as a cognitive agent of human intentionality, will allow humans to externalize violence in ways that are removed in space and time and thus reconfigure established notions of defense, deterrence, and diplomacy.

While phenotypic expressions and active externalism helps explain how VIS are *expressions* of and *aids* to human violence, they do not explain how networks of human violence distribute responsibilities between human and non-human actants.

---

[489] Andy Clark and David Chalmers, *The Extended Mind*, 58(1) ANALYSIS, 8 (1998).

[490] Clark and Chalmers, *supra* note 489 at 9.

Actor Network Theory ("ANT") can help explain how a violence network operates without undermining fundamental rules of legal responsibility. ANT is a descriptive tool that helps tell "stories about how relations assemble or don't."[491] It is a form of material-semiotics that seeks to advance a relational materiality where "all entities achieve significance in relation to others."[492] At the heart of the ANT framework is the notion that human and non-human "actants" operate to make up networks.

In *The Pasteurization of France*, Bruno Latour, a leading proponent of ANT, uses the process of scientific discoveries to illustrate how the actor-networks operates. While scientists, such as Louis Pasteur, are "actors" known for driving discoveries in laboratories, we cannot speak of "science" without also speaking of other "actants", such as microbes, laws, microscopes, research groups, and research institutions. Actors and actants give meaning to each other through their relationships in the network. Importantly, human actors act as spokespersons in the assemblage of that network. Spokespersons mobilize actants in the network to achieve certain goals. Thus, "Pasteur speaks for microbes, the Curies can be said to speak for plutonium, Cantor for transfinite numbers, Einstein for photons."[493] ANT helps explain how "collective actions" are performed in networks where actors (humans) and actants

---

[491] John Law, *Actor-network theory and material semiotics*, in BRYAN S. TURNER, THE NEW BLACKWELL COMPANION TO SOCIAL THEORY141-142 (Blackwell Publishing, 3rd ed., 2008).

[492] Cassandra S. Crawford, *Actor Network Theory*, ENCYCLOPEDIA OF SOCIAL THEORY, *available at* http://sk.sagepub.com/reference/socialtheory/n1.xml.

[493] Sal Restivo, *Bruno Latour: The Once and Future Philosopher*, in GEORGE RITZER AND JEFFREY STEPINSKY (eds.), THE NEW BLACKWELL COMPANION TO MAJOR SOCIAL THEORISTS 538-539 (Blackwell, 2011).

(non-humans) share practical (but not legal) responsibilities,[494] and without anthropomorphizing actants.[495]

ANT frameworks can be used where there is semiotic relationality (a network whose components define and shape one another), heterogeneity (diversity of actors, human and otherwise), process (components need to play their respective parts), and space and scale (networks extend themselves and translate distant actors).[496] In the context of human violence, the military is a network of violence that includes a multitude of actants, including (in no particular order) political leaders, policy makers, military lawyers, commanders, soldiers, contractors, researchers, victims, fuel, weapons, weapons systems, ammunition, computers, and code. The VIS is an actant whose meaning depends on its relations to other components in the network. Human beings, including political leaders, programmers, commanders, and soldiers, are actors that shape and define those relations. In effect, human beings are the "spokespersons" of the means and methods of human violence (e.g., VIS) just as Pasteur was the "spokesperson" for the microbes he studied and the pasteurization process he pioneered.

Given their anthropocentric bent, systems of law continuously strive to identify human beings and human institutions as "spokespersons" of human relations

[494] These actor-networks are often "punctualized", meaning that the components of complex systems are hidden from the view of the user or audience of the system. Thus, while human beings may be "hidden from the view" of the operation of VIS, they remain the most important "components" in the "complex" system of human violence. *See* Restivo, *supra* note 493, at 541-542.

[495] Restivo, *supra* note 493, at 539 (the actant (microbes, photon, the laboratory) are the non-anthropomorphic sibling of "actor" and plays an important part in the network).

[496] Law, *supra* note 491, at 146.

with non-human actants.[497] Chapter III demonstrated how human institutions and individuals remain responsible for the use of VIS. Attempts to argue that VIS can "speak" for themselves contradicts the fundamental structures of law and violence: humans, as the true subjects of the law, ultimately remain responsible for the violence they inflict. Accordingly, regulating the use of force in international armed conflicts is necessarily an exercise in parsing through networks of human violence to determine the human locus in emerging violent technologies.[498] By focusing on the idea that human violence is performed by humans and non-human actants, this reconceptualization can assist "courts, tribunals and other agents" to continue expanding rights and responsibilities "to encompass wider circles of conduct, and *additional actors within conflicts*."[499]

Attitudes toward VIS in international armed conflicts will change as social expectations and cultural priorities shift to welcome the increasing digitization of human activity. As with all forms of human violence, the meaning of VIS will be shaped not only by how they are built but also by how they are semiotically and culturally construed.[500] Culture will cultivate the meanings underlying the

---

[497] The non-human is not foreign to law. Indeed, we accept evidence from scientific analysis (DNA analysis) for example as part of a trial. Latour characterized this as "the testimony of nonhumans". BRUNO LATOUR, WE HAVE NEVER BEEN MODERN 23 (Harvard University Press, 1993) ("These nonhumans, lacking souls but endowed with meaning, are even more reliable than ordinary mortals, to whom will is attributed but who lack the capacity to indicate phenomena in a reliable way.").

[498] TEITEL, *supra* note 207, at 4 ("In interpreting and elaborating the law of humanity, courts, tribunals, and other agents have had to address tensions between, and gaps within, the different traditional doctrinal sources of humanity law.").

[499] *Id.*, at 4 (emphasis added).

[500] As Geertz put it: "[b]elieving, with Max Weber, that man is an animal suspended in webs of significance he himself has spun, I take culture to be those webs, and the analysis of it to be

technology, including what they are (combatants, human-like machines, weapons?) and what they are for (to promote peace, protect hospitals, commit genocide?).[501] As Wittgenstein put it, "[t]he arrow points only in the application that a living being makes of it."[502] Meanings are subject to a matrix of semiotic constructions that adapt to changing circumstances.[503] If used to *stop* or *prevent* war crimes, the technologies underlying VIS are viewed favorably and very differently from, say, if a terrorist group uses them to commit a terrorist attack.

Different cultures may internalize technological shifts differently. For example, while "the threat of humanity creating things [it] cannot control pervades Western literature"[504] this "fear does not pervade Eastern culture to the same extent."[505] As Alex Ross notes in *The Industries of the Future*:

> The cultural dynamic in Japan is representative of the culture through much of East Asia, enabling the Asian robotics industry to speed ahead, unencumbered by cultural baggage. Investment in robots reflects a cultural comfort with robots, and, in China, departments of automation are well represented and well respected in the academy. There are more than 100 automation departments in Chinese

---

therefore not an experimental science in search of law but an interpretive one in search of meaning." CLIFFORD GEERTZ, THE INTERPRETATION OF CULTURES, SELECTED ESSAYS 5 (Basic Books, 1973).

[501] *See* Brehm, *supra* note 74, at 4 ("Meaning-making is a social practice by which human beings interact with each other to make *common sense* of the world. Meaning includes moral understandings of right and wrong, cognitive understandings of true and false, perceptual understandings of like and unlike.").

[502] *Ludwig Wittgenstein (1889-1951)*, INTERNET ENCYCLOPEDIA OF PHILOSOPHY, *available at* http://www.iep.utm.edu/wittgens/#H5.

[503] MARKOFF, *supra* note 62, at xviii (noting "the social construction of technology—the understanding that we shape our tools rather than being shaped by them.").

[504] ROSS, *supra* note 327, at 23.

[505] *Id.* ("Prometheus was condemned to an eternity of punishment for giving fire to humans. When Icarus flew too high, the sun melted his ingenious waxed wings and he fell to his death. In Mary Shelley's *Frankenstein*, Dr. Frankenstein's grotesque creation wreaks havoc and ultimately leads to its creator's death and numerous B-movie remakes.").

> universities, compared with approximately 76 in the United States despite the larger total number of universities in the United States. In South Korea, teaching robots are seen in a positive light; in Europe, they are viewed negatively. As with eldercare, in Europe robots are seen as machines, whereas in Asia they are viewed as potential companions.[506]

As technology becomes increasingly embedded in our daily lives, workplaces, and relationships,[507] cultural expectations are likely to further anthropomorphize intelligent systems.[508] VIS could in the future be described as "soldiers," "fighters," or "combatants."

This perceptive shift, however, should not affect the social expectation that humans remain responsible for the use of VIS.[509] We should not let the anthropomorphization of technology undermine the application of law and the allocation of responsibility—particularly for technologies that are capable of inflicting violence. That may of course change if VIS and robots generally start

---

[506] *Id.*, at 23.

[507] Heyns, *supra* note 242, in BHUTA, BECK, GEISS, LIU AND KRESS, *supra* note 30, 3 ("Computers affect almost all aspects of our lives and have become an integral part not only of our world but also of our very identity as human beings.").

[508] As Martin Ford notes in the context of the workplace, "machines themselves are turning into workers." FORD, *supra* note 321, at xii.

[509] Of course, it is difficult to predict the social and cultural expectations of the future. *See* Johnson, *supra* note 14, at 8 ("Recognizing that responsibility is embedded in relationships adds further support to the idea that the nature of a particular technology does not necessitate a particular responsibility arrangement. Accountability relationships are not dictated by nature or anything else. The nature of a technology is relevant to the responsibility arrangements, but responsibility arrangements are socially constituted through the norms and expectations of particular activities and contexts. . . . Hence, when it comes to responsibility-accountability for artificial agents of the future, the possibilities are open. People of the future might accept no human responsibility; they might come to expect robots to explain their behavior specifying why they did what they did; they might hold robot manufacturers accountable or they might hold multiple parties accountability according to their particular contributions to robot behavior.").

acquiring legal personhood,[510] or as Arendt put it, "right to have rights."[511] But since VIS will always act within networks of human violence (unless they figure out a way to procreate and be energy independent!), humans, as creators and users of these systems, should always remain responsible for consequences of their use.[512]

---

[510] Upendra Baxi, *The Posthuman and Human Rights*, in UPENDRA BAXI, HUMAN RIGHTS IN A POSTHUMAN WORLD 198-199 (Oxford University Press, 2009) ("Personification of the non-human has a very long juridical history, although it is with more recent invention of citizenship and corporate personality that the distinction between 'natural' and 'artificial' person (natural persons are born and the artificial persons are made) becomes problematic. The distinction misleads because in both cases it is the law that assigns personhood. And citizenship is an artificial personhood just the same as the corporation; not all born humans 'enjoy' the 'blessing' of citizenship-being. Stateless, diasporic, and nomadic humans may not belong to the category of persons/ populaces and may be reduced to status of things and objects, denied in Arendt's immortal phrase 'the right to have rights.'"); *see also* United States v. Certain Real Prop. & Premises Known as 38 Whalers Cove Drive, Babylon, N.Y., 747 F. Supp. 173, 177 (E.D.N.Y. 1990), aff'd, 954 F.2d 29 (2d Cir. 1992) ("the ascription of personality to offending objects persisted into the modern law of civil forfeiture.").

[511] HANNAH ARENDT, THE ORIGINS OF TOTALITARIANISM 294 (Harcourt, Brace & Co., 1951).

[512] Whether in the future that responsibility becomes shared is a possibility, though the point is that humans should never be conceived as being removed from the responsibility chain.

Chapter V

Conclusion

This thesis has sought to challenge the perspective that lethal weapons capable of autonomous operation ought to be pre-emptively banned on the grounds that they will be incapable to satisfy IHL and will allow human beings to evade liability for their use. It has argued that such dystopian prognoses about VIS are based on a pessimistic and outdated view of artificial intelligence, and moreover, on a misinterpretation of international law's ability to regulate means and methods of human violence. In the process, this thesis has also critiqued terminologies of "lethal autonomy" and sought to refocus the debate toward violence, intelligence, and weapon systems. In the final analysis, an alternative narrative emerges—one that recognizes technology's potential to better bridge the gap between "norms and reality." Indeed, advances in machine learning and robotics will allow, at minimum, for rigorous adherence to IHL, and at best, that VIS outperform human compliance with IHL on the battlefield.

As AI becomes more influential in all corners of human life,[513] decision-making will increasingly be made or shaped by powerful algorithms both on and off the battlefield.[514] But human beings cannot cede their responsibility for violence by

[513] *See* MARKOFF, *supra* note 62, at xi ("Robots are pervading our daily lives. Cheap sensors, powerful computers, and artificial intelligence software will ensure they will, increasingly, be autonomous. They will assist us and they will replace us."); *see* also *id.*, at xv-xvi ("With the arrival of 'ubiquitous computing,' . . . smart machines are making decisions for us.").

[514] Lewis, Blum, and Modirzadeh, *supra* note 6, at i ("[A]uthority is increasingly expressed algorithmically. War is no exception. . . . Warring parties have long expressed authority and power

merely delegating functions to algorithms. Human violence has always been expressed in networks of human and non-human actants that collaborate to implement violent political[515] objectives.[516] While actants can share practical responsibilities in these networks, there is no correlated distribution of legal responsibilities from humans to other actants. Indeed, human institutions and human beings remain responsible for the use of VIS under public international law. Accordingly, VIS should not be pre-emptively banned.

This conclusion should not be taken as a justification for the increased militarization of States or other actors. To the contrary, it serves as a warning to States and human agents that decisions to develop and deploy VIS carry responsibilities that cannot be discarded by virtue of the technology's complexity. In many ways, this heightens the bar to responsible development, deters poor deployment, and increases the expectations that VIS must categorically comply with IHL. Additional accountability methods will also need to be developed beyond the

---

through algorithms. For decades, algorithms have helped weapons systems—first at sea and later on land—to identify and intercept inbound missiles. Today, military systems are increasingly capable of navigating novel environments and surveilling faraway populations, as well as identifying targets, estimating harm, and launching direct attacks—all with fewer humans at the switch.") (internal citation omitted).

[515] War, as a mere continuation of human-made policy "by other means", is an "act of violence" that works as a "political instrument" to achieve political goals. CLAUSEWITZ, *supra* note 182, at 5, 22.

[516] Some have sought to characterize this collaborative network as a post-human distribution of human subjectivity and agency, re-situating "the human as an intelligent machine in constant interaction with other intelligent machines." In this post-human context, distributed cognition between human and non-human intelligence replaces the notion of autonomous will. This partnership between humans and non-human actants "is not so much a usurpation of human right and [human] responsibility as it is a further development in the construction of distributed cognition environments". As technology becomes more deeply embedded in the human condition, the notion of what is 'human' be stretched to allow other actants to participate in the mediation of human conduct. Indeed, in potentially allowing for greater adherence to IHL, VIS may evolve to become "constitutive partners" in the development of how human rights are shaped, promoted and protected. Baxi, *supra* note 510, at 205, 223.

legal architecture.[517] For example, Lewis, Blum and Modirzadeh identify a number of "soft mechanisms" to strengthen accountability for the use of weapons like VIS. These mechanisms include independent monitoring, non-binding declarations and resolutions, normative design and technical architectures, and community self-regulation.[518]

Ultimately, stakeholders should recognize that bellicosity does not spring from the human hand but is seeded by the human mind. Code, as a functional extension of human intentionality, extends human action in space and time in ways that challenge established notions of the structure of the international order. The use of drones and cyberwarfare are already undermining traditional state-centric concepts, such as sovereignty, territoriality, and the monopoly of force. It is only a matter of time before VIS upend the organization of warfare.

To this end, the weaponization of artificial intelligence[519] may shift power from entities that monopolize force (traditionally States) to those that monopolize or control information technology. Code will become to the 21st century what the nuclear weapon was to the 20th. But since coding is available to anyone and technologies are becoming more accessible and cheaper to acquire, the greatest

[517] Lewis, Blum, and Modirzadeh, *supra* note 6, at xii ("In short, individual responsibility for international crimes under international law remains one of the vital accountability avenues in existence today, as do measures of remedy for state responsibility. Yet in practice responsibility along either avenue is unfortunately relatively rare. And thus neither path, on its own or in combination, seems to be sufficient to effectively address the myriad regulatory concerns pertaining to war algorithms—at least not until we better understand what is at issue. These concerns might lead those seeking to strengthen accountability of war algorithms to pursue not only traditional, formal avenues but also less formal, softer mechanisms.").

[518] Lewis, Blum, and Modirzadeh, *supra* note 6, at 91.

[519] ROSS, *supra* note 327, at 121 ("The weaponization of code").

defense challenge of the 21st century will be to maximize human security in an age where vulnerability and defense are distributed—or democratized—among public and private actors.[520] As Wittes and Blum note, "[i]n our new world, you can pose a threat to the security of every state or person on the planet—and each can also threaten you."[521] In this new world, human impunity is not an option. Much like a manufacturer is responsible for the products she makes—regardless of whether the assembly line is automated—humans, as manufacturers of violence, are responsible for its effects.

While the future of violence will reducibly be waged through code, its malleability and openness allows it to also work as a medium through which law regulates conduct. Indeed, code may allow for a more rigorous translation of IHL principles into military operation. Self-learning code, coupled with powerful robotics, will allow for greater precision, discrimination, and proportionality in attacks. It may also encourage the use of non-lethal force. And unlike human intelligence, code is testable, scalable, upgradeable, and connectable. As Laplace put it in another context, which is nonetheless pertinent here:

> An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes.[522]

---

[520] WITTES AND BLUM, *supra* note 2, at 2 ("[D]estructive power once reserved to states is now the potential province of individuals."); MARKOFF, *supra* note 62, at xvii ("More than merely replacing humans, information technology is democratizing certain experiences.").

[521] WITTES AND BLUM, *supra* note 2, at 6.

[522] PIERRE SIMON LAPLACE, A PHILOSOPHICAL ESSAY ON PROBABILITIES 282, translated by Frederick Wilson Truscott and Frederick Lincoln Emory, (Dover Publications, 1951).

In addition to creating new avenues to produce violence, artificial intelligence may also help humans find optimal solutions to international conflicts that wholly avoid the use of force. The dialectical struggle between the weaponization of artificial intelligence and the human quest for global peace may refurbish our perspective of AI as a tool built not to destroy Mankind, but to save it from itself.[523] But before we find the algorithm for peace, our species still has considerable work to do in humanizing the inconvenience of war.

---

[523] MARKOFF, *supra* note 62, at xix (noting Nortbert Wiener's warning about the potential of automation: "We can be humble and live a good life with the aid of machines, or we can be arrogant and die.").

Bibliography

"Active Denial Technology Fact Sheet." U.S. Department of Defense, Non-Lethal Weapons Program (May 11, 2016), http://jnlwp.defense.gov/Press-Room/Fact-Sheets/Article-View-Factsheets/Article/577989/active-denial-technology-fact-sheet/.

Allenby, Brad. "What Human Rights Watch's "Case Against Killer Robots" Gets Wrong About Military Realities." *SLATE* (Nov. 20, 2012), http://www.slate.com/blogs/futuretense/2012/11/20/human rightswatch_s_case_against killer robotsreport misunderstands _the.htm.

Allhoff, Fritz, Nicholas G. Evans, Adam Henschke (eds.), *Routledge Handbook of Ethics and War*. New York: Routledge, 2014.

Anderson, Kenneth, Daniel Reisner and Matthew Waxman. "Adapting the Law of Armed Conflict to Autonomous Weapon Systems." *Int'l L. Stud.* 90 (2014): 386.

Anderson, Kenneth, and Matthew Waxman. "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can." *Hoover Institution, Research Paper* (2013).

Arendt, Hannah. *The Origins of Totalitarianism*. New York: Harcourt, Brace & Co., 1951.

Arkin, Ronald. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL: Chapman and Hall, 2009.

Arkin, Ronald, "Lethal Autonomous Systems and the Plight of the Non-Combatant." *AISB Quarterly*, no. 137 (July 2013).

Arkin, Ronald. "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, Part I: Motivation and Philosophy." Georgia Institute of Technology (2007), http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

Articles on the Responsibility of States for Intentionally Wrongful Acts, in Report of the International Law Commission to the General Assembly on Its Fifty-Third Session, 56 U.N. GAOR Supp. No. 10, at 1, 43, U.N. Doc. A/56/10 (2001)

Asaro, Peter. "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making." *International Review of the Red Cross* 94, no. 886 (2012).

Baxi, Upendra, *Human Rights in a Posthuman World*. New Dehli: Oxford University Press, 2009.

Beard, Jack, "Autonomous Weapons and Human Responsibilities." *Georgetown Journal of International Law* 45 (2014): 617.

Bell, Jason, *Machine Learning: Hands On For Developers and Technical Professionals*. Indianapolis, IN: Wiley, 2015.

Bento, Lucas, "Could Science Defeat Terrorism? Using Robots to Hunt Down ISIS." *The Diplomat*, August 31, 2015.

Bhuta, Nehal, Susanne Beck, Robin Geiss, Hin-Yan Liu and Claus Kress, *Autonomous Weapons Systems: Law, Ethics, and Policy*. Cambridge, UK: Cambridge University Press, 2016.

Blum, Gabriella, "The Dispensable Lives of Soldiers." *Journal of Legal Analysis* 2(1) (2010): 69.

Brehm, Maya, "Meaningful Human Control." Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 14, 2015).

Brooks, Rosa. *How Everything Became War and the Military Became Everything.* New York: Simon & Schuster, 2016.

Brooks, Rodney. "Artificial intelligence is a tool, not a threat." *Rethink Robotics Blog*, November 10, 2014, http://www.rethinkrobotics.com/artificial-intelligence-tool-threat/

Brynjolfsson, Erik, and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company, 2016.

Cassese, Antonio. *International Criminal Law*. Oxford: Oxford University Press, 2008.

"Cadre Juridique D'un Eventuel Developpement et Usage Operationnel d'un Futur Systeme D'armes Letal Autonome (Sala)." Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 2016).

"Canadian Food for Thought Paper: Mapping Autonomy." Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW), 2006, http://www.unog.ch/

80256EDD006B8954/(httpAssets)/C3EFCE5F7BA8613BC1257F8500439B9F/$file/2016_LAWS+MX_CountryPaper+Canada+FFTP1.pdf.

Caton, Jeffrey L., “Autonomous Weapon Systems: A Brief Survey of Developmental Operational, Legal, and Ethical Issues.” The Letort Papers, United States Army War College Press, 2015.

“The Chip That’s Bad at Math.” *MIT Technology Review* 119 (2016): 4.

Chopra, S., & White, L. W. *A Legal Theory For Autonomous Artificial Agents*. Michigan: The University of Michigan Press, 2011).

Clark, Andy, and David Chalmers. “The Extended Mind.” *Analysis* 58(1) (1998): 8.

Clark, Liat, “Google’s Artificial Brain Learns to Find Cat Videos.” *Wired,* June 26, 2012, https://www.wired.com/2012/06/google-x-neural-network/.

Clausewitz, Carl von, *On War*. London: Wordsworth Editions, 2000.

Coke, Sir Edward. *Institutes of the Laws of England*. Volume III 58 (Printed for A. Crooke [and 12 others], 1669).

Crawford, Cassandra S., “Actor Network Theory.” *Encyclopedia of Social Theory*, http://sk.sagepub.com/reference/socialtheory/n1.xml.

Crawford, James. *State Responsibility: The General Part*. Cambridge, UK: Cambridge University Press, 2014.

Crowe, Jonathan, and Kylie Weston-Scheuber. *Principles of International Humanitarian Law*. Northampton, MA: Edward Elgar, 2013.

Davenport, Christian. “Robots, swarming drones and ‘Iron Man’: Welcome to the new arms race.” *The Washington Post*, June 17, 2016, https://www.washingtonpost.com/news/checkpoint/wp/2016/06/17/robots-swarming-drones-and-iron-man-welcome-to-the-new-arms-race/

Dawkins, Richard. *The Extended Phenotype*. Oxford: Oxford University Press, 1999.

DeLanda, Manuel. *War in the Age of Intelligent Machines*. New York: Zone Books, 1991.

Demy, Timothy J., and George R. Lucas (eds.). *Military Ethics and Emerging Technologies*. New York: Routledge, 2014.

DeYoung, Karen, and Greg Miller, “White House releases its count of civilian deaths in counterterrorism operations under Obama.” *Washington Post*, July 1, 2016, https://www.washingtonpost.com/world/national-security/white-house-releases-its-count-of-civilian-deaths-in-counterterrorism-operations-under-obama/2016/07/01/3196aa1e-3fa2-11e6-80bc-d06711fd2125_story.html.

Directive Number 3000.09, U.S. Department of Defense, November 21, 2012, http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf

Domingos, Pedro. *The Master Algorithm*. New York: Basic Books, 2015.

Dumouchel, Paul, and Luisa Damiano. *Vivre Avec Les Robots*. Paris: Seuil, 2016.

Dunne, Philip, "UK and US look to robotics help for 'last mile.'" U.*K. Ministry of Defense News*, July 14, 2016, http://www.gov.uk/government/news/uk-and-us-look-to-robotics-help-for-last-mile

Dunsany, Lord. *The Last Revolution*. New York Talos, reprint edition, 2015.

Eitel, Andreas, Jost Tobias, Springenberg, Luciano Spinello, Martin Riedmiller, and Wolfram Burgard. "Multimodal Deep Learning for Robust RGB-D Object Recognition." *arXiv*, August 18, 2015, http://www2.informatik.uni-freiburg.de/~spinello/eitelIROS15.pdf.

"Entrenching Impunity Government Responsibility for International Crimes in Darfur." *Human Rights Watch*, https://www.hrw.org/sites/default/files/features/darfur/fiveyearson/report9.html

Espada, Cesáreo Gutiérrez and María José Cervell Hortal. "Autonomous Weapons Systems, Drones, and International Law." *Revista del Instituto Español de Estudios Estratégicos* 2 (2013): 4.

Flatow, Ira. "Science Diction: The Origin of the Word 'Robot'." *NPR*, April 22, 2011, http://www.npr.org/2011/04/22/135634400/science-diction-the-origin-of-the-word-robot

Ford, Martin. *Rise of the Robots: Technologgy and the Threat of a Jobless Future*. New York: Basic Books, 2015.

Frankish, Keith and William M Ramsey (eds.). *The Cambridge Handbook of Artificial Intelligence.* Cambridge: Cambridge University Press, 2014.

Gady, Franz-Stefan. "Meet Russia's New Killer Robot." *The Diplomat*, July 21, 2015, http://thediplomat.com/2015/07/meet-russias-new-killer-robot/.

Geertz, Clifford. *The Interpretation of Cultures, Selected Essays*. New York: Basic Books, 1973.

Gravett, Chris. *Tudor Knight*. Oxford: Osprey Publishing, 2006.

Greenwald, Ted . "Does Artificial Intelligence Pose a Threat." *Wall Street Journal*, May 10, 2015), http://www.wsj.com/articles/does-artificial-intelligence-pose-a-threat-1431109025

Hart, H.L.A. *Punishment and Responsibility*. Oxford: Oxford University Press, 2nd edition, 2008.

Henckaerts, Jean-Marie, and Louise Doswald-Beck (eds.). *Customary International Humanitarian Law*. Cambridge: Cambridge University Press, 2005.

Henn, Fritz, Norman Sartorius, Hanfried Helmchen, and Hans Lauter. *Contemporary Psychiatry*. München, Germany: Springer, 2001.

Horowitz, Michael C. and Paul Scharre. "Meaningful Human Control in Weapon Systems: A Primer." Working Paper, Center for New American Security, March 2015, http://www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_Working_Paper_031315.pdf.

Hayden, Michael V. "To Keep America Safe, Embrace Drone Warfare." *New York Times*, February 19, 2016, http://www.nytimes.com/2016/02/21/opinion/sunday/drone-warfare-precise-effective-imperfect.html?_r=0

Hellstrom, T. "On the moral responsibility of military robots." *Ethics and Information Technology* 5 (2013): 99.

Heyns, Christof. "Report of the Special Rapporteur on extrajudicial, summary or arbitrary Executions." Human Rights Council, 3rd Session, A/HRC/23/47 (Apr. 9, 2013).

Holst, P. K., and P. Funk (eds.). *Frontiers in Artificial Intelligence and Applications.* Netherlands: IOS Press Books, 2008).

Geiss, Robin, *Autonomous Weapons Systems: Risk Management and State Responsibility*, Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW) (Apr. 2016) http://www.unog.ch/80256EDD006B8954/(httpAssets)/00C95F16D6FC38E4C1257F9D0039B84D/$file/Geiss-CCW-Website.pdf

Gentner D., Holoyoak, K.J., and Kokinov, B..N. (eds.). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: MIT Press, 2001).

International Covenant on Civil and Political Rights, G.A. Res. 2200A, U.N. GAOR, 21st Sess., Supp. No. 16, U.N. Doc. A/6316 (1966), 999 U.N.T.S. 171

Isaacson, Walter. “The Intersection of the Humanities and the Sciences.” Lecture, National Endowment for the Humanities (2014), https://www.neh.gov/about/awards/jefferson-lecture/walter-isaacson-lecture.

Jasinski, Michael P. *Social Trust, Anarchy, and International Conflict*. New York: Palgrave Macmillan, 2011.

Johnson, Deborah G. “Technology with No Human Responsibility?” *Journal of Business Ethics* 127(4) (2014): 707.

Jordan, Bryant, *Pentagon Wants Artificial Intelligence to Defeat Enemy Networks*, Defense Tech (May 5, 2016), http://www.defensetech.org/2016/05/05/pentagon-wants-artificial-intelligence-to-defeat-enemy-networks/.

Kanwar, Vik, *Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons*, 2 Harvard National Security Journal 616, 618 (2011)

Kastan, Benjamin. “Autonomous Weapons Systems: A Coming Legal ‘Singularity’?” *Journal of Law, Technology, & Policy* 1 (2013): 45.

Kiai, Maina and Christof Heyns. “Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies.” United Nations General Assembly, Human Rights Council, A/HR/31/66, at 15 (Feb. 4, 2016)

“Killer Robots: UK Government Policy on Fully Autonomous Weapons.” Article 36, April, 2013, http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf

Kahneman, Daniel. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

Koza, John R. “Genetic Programming.” *Genetic Programming Website*, http://geneticprogramming.com/tutorial/

Krishnan, Armin. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. London: Routledge, 2009.

Kurzweil, Ray. “MIT deep-learning system autonomously learns to identify objects.” *Kurzweil Accelerating Intelligence*, May 14, 2016, http://www.kurzweilai.net/mit-deep-learning-system-autonomously-learns-to-identify-objects

Kurzweil, Ray. *The Singularity is Near*. New York: Penguin, 2006.

Krizhevsky, Alex, Ilya Sutskever, Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” Neural Information ProcessingSystems Foundation (2012), https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Lample, Guillaume, and Devendra Singh Chaplot. “Playing FPS Games with Deep Reinforcement Learning.” *arXiv*, September 2016, https://arxiv.org/abs/1609.05521

Laplace, Pierre Simon. *A Philosophical Essay on Probabilities*. Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: Dover Publications, 1951.

Latour, Bruno. “*On actor network theory: a few clarifications*.” *Soziale Welt* 47 (1996): 360.

Latour, Bruno. *We Have Never Been Modern*. Cambrige, MA: Harvard University Press, 1993.

“Law of War Manual.” U.S. Department of Defense, June, 2015.

Le, Quoc V., Marc’Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeff Dean, Andrew Y. Ng. “Building High-level Features Using Large Scale Unsupervised Learning.” Research Google (2012), http://static.googleusercontent.com/media/research.google.com/en//archive/unsupervised_icml2012.pdf

Leuthy, Cameron. “In battle, you need to trust your robots.” *Bloomberg Government*, June 16, 2016, http://about.bgov.com/blog/battle-need-trust-robots/.

Lieber, Francis, *Instructions for the Government of Armies of the United States in the Field*. New York: D. Van Nostrand, 1863.

Lewis, Dustin A., Gabriella Blum, and Naz K. Modirzadeh, *War-Algorithm Accountability*. Harvard Law School Program on International Law and Armed Conflict, 2016.

“Losing Humanity: The Case against Killer Robots.” Human Rights Watch, November 2012.

“Ludwig Wittgenstein (1889-1951).” *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/wittgens/#H5.

Marchant, Gary E., Braden Allenby, Ronald Arkin, Edward T. Barrett, Jason Borenstein, Lyn M. Gaudet, Orde Kittrie, Patrick Lin, George R. Lucas, Richard O’Meara, Jared Silberman. “International Governance of

Autonomous Military Robots." *The Columbia Science and Technology Law Review*, 17 (2011): 277.

Markoff, John. *Machines of Loving Grace*. HarperCollins, 2015.

Mariani, Joe, Brian Williams, and Brett Loubert. "Continuing the march: the past, present, and future of the IoT in the militar*y*." Deloitte University Press, August 6, 2015), http://dupress.com/articles/internet-of-things-iot-in-military-defense-industry/.

Matthias, A. "The responsibility gap: Ascribing responsibility for the actions of learning Automata." *Ethics and Information Technology* 6(3) (2004):175–183.

May, L. and Edenberg, E., (eds.). "Jus post bellum" and transitional justice. Cambridge: Cambridge University Press, 2013.

Mitchell, Tom M. *Machine Learning*. New York: McGraw-Hill, 1997.

Murphy, Robin, and David Woods. "Beyond Asimov: The Three Laws of Responsible Robotics." *IEEE Intelligent Systems*, July-August 2009, http://ts-si.org/files/IEEEIS_WebExtra-0709.pdf

Meron, Theodor. "The Humanization of Humanitarian Law." *The American Journal of Intenrational Law* 94(2) (2000): 239.

Marino, D., & Tamburrini, G. "Learning robots and human responsibility." *International Review of Information Ethics* 6 (2006): 46–51.

McCormick, Ty. "Lethal Autonomy: A Short History." *Foreign Policy (Online)*, January 24, 2014, http://foreignpolicy.com/2014/01/24/lethal-autonomy-a-short-history/.

McFarland, Tim, and Tim McCormack. "Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes?" Int'l L. Stud. 90 (2014): 361.

Melzer, Nils. "Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare." European Parliament, Directorate-General for External Policies, May 2013.

"Mind the Gap: The Lack of Accountability for Killer Robots." Human Rights Watch (2015).

Mindell, David A. *Our Robots, Ourselves*. New York: Viking, 2015.

Moran, M.E. "*The da Vinci robot*." *Journal of Endourology* 20(12) (2006): 986–990.

Nagenborg, M., Capurro, R., Weber, J., & Pingel, C. "Ethical regulations on robotics in Europe." *AI & Society* 22 (2008): 349–366.

Noorman, Merel, and Deborah G. Johnson. "Negotiating autonomy and responsibility in military robots." *Ethics and Information Technology* 16 (2014): 51.

Oeter, Stefan (ed.). *The Handbook of International Humanitarian Law*. Oxford: Oxford University Press, 3d edition, 2013.

"Off Target. The Conduct of the War and Civilian Casualties in Iraq." Human Rights Watch at 19 (2003), http://www.hrw.org/reports/2003/12/11/target.

Osborn, Kris. "DARPA Unveils Hack-Proof Drone, Defense Tech." October 3, 2016, http://www.defensetech.org/2014/05/21/darpa-unveils-hack-proof-drone/.

Pagallo, Ugo. "Robots of just war: a legal perspective." *Philosophy and Technology* 24 (2011): 307– 323.

Parker, Geoffrey (ed.). *The Cambridge History of Warfare*. Cambridge: Cambridge University Press, 2005.

Pellerin, Cheryl. "Work: Human-Machine Teaming Represents Defense Technology Future." *U.S. Department of Defense News*, November 8, 2015, http://www.defense.gov/News/Article/Article/628154/work-human-machine-teaming-represents-defense-technology-future

Pervukhin, Anna. "Deodands: A Study in the Creation of Common Law Rules." *The American Journal of Legal History* 47(3) (2005): 237.

Pinker, Steven. *The Better Angels of Our Nature*. New York: Penguin Books, 2011.

Rieland, Randy. "Can Killer Robots Learn to Follow the Rules of War?" *Smithsonian Magazine*, May 29, 2014.

Sassóli, Marco. "Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified." *Int'L. L. Stud.* 90 (2014): 308.

Schmitt, Michael N., and Jeffrey S. Thurnher. "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict." *Harvard National Security Journal* 4 (2013): 231.

Rawlins, Aimee. "Math is racist: How data is driving inequality." *CNN Money*, September 6, 2016), http://money.cnn.com/2016/09/06/technology/weapons-of-math-destruction/index.html.

Republic of Sierra Leone, Opening Statement by Ambassador Yvette Stevens, CCW 2016 Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), April 2016, http://www.unog.ch/80256EDD006B8954/(httpAssets)/0054AE2FAA24E566C1257F9B004A2CAB/$file/SIERRA+LEONE+GENERAL+STATEMENT+2016+MEETING+ON+LAWS.pdf

Ritzer, George, and Jeffrey Stepinsky (eds.). *The New Blackwell Companion to Major Social Theorists*. Oxford: Blackwell, 2011.

Ross, Alec. *The Industries of the Future*. New York: Simon & Schuster, 2016.

Roudgarmi, P., Monavari, M., Feghhi, J. et al. "Environmental impact prediction using remote sensing images." *Journal of Zhejiang University-Science A* 9(3) (2008): 381-390.

Rousseau, Jean-Jacques. *On The Social Contract*. New York: Dover Thrift Edition, 2003.

Roscini, Marco. *Cyber Operations and the Use of Force in International Law*. Oxford: Oxford University Press, 2014.

"The Role of Autonomy in DoD Systems." Task Force Report, United States Department of Defense, Defense Science Board, July 2012, http://fas.org/irp/agency/dod/dsb/autonomy.pdf.

Russell, Stuart, Daniel Dewey, Max Tegmark. "Research Priorities for Robust and Beneficial Artificial Intelligence." *Future of Life*, Winter 2015, http://futureoflife.org/data/documents/research_priorities.pdf at 107-108

Sarat, Austin, Lawrence Douglas, and Martha Merrill Umphrey. *Law and War*. Stanford: Stanford University Press, 2014.

Saxon, Dan, "In the Context of the Design and Use of Autonomous Weapon Systems, What is Judgment?" Presentation, Informal Meeting of Experts on Lethal Autonomous Weapons Systems of the Convention on Certain Conventional Weapons (CCW), April 2016.

Schmitt, Michael N. "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics." *Harv. Nat'l. Sec. J. Features* (2013), http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf

Sharkey, Noel E. "The evitability of autonomous robot warfare." *Internatioanl Review of the Red Cross*, 94, no. 886, (2012): 789.

Spitz, Francois, and Sovan Lek. "Environmental impact prediction using neural network modelling. An example in wildlife damage available." *Journal of Applied Ecology* 36(1) (April 1999): 317.

Socher, Richard, Brody Huval, Bharath Bhat, Christopher D. Manning, Andrew Y. Ng. "Convolutional-Recursive Deep Learning for 3D Object Classification." Neural Information Processing Systems Foundation (2012), http://nlp.stanford.edu/pubs/SocherHuvalBhatManningNg_NIPS2012.pdf

Sugiyama, Masahi. *Introduction to Statistical Machine Learning*. Waltham, MA: Morgan Kaufman, 2016.

Sparrow, R. "Killer robots." *Journal of applied philosophy* 41(1) (2007): 62–77.

Teitel, Ruti G. *Humanity's Law*. New York: Oxford University Press, 2011.

Turner, Bryan S. *The New Blackwell Companion to Social Theory*. Oxford: Blackwell Publishing, 3rd ed., 2008.

Thorne, Terri, and Eric C. Husby. "Law of Armed Conflict Deskbook." The Judge Advocate General's Legal Center and School, U.S. Army, 2012.

Tucker, Patrick. "The Pentagon is Nervous about Russian and Chinese Killer Robots." *Defense One*, December 14, 2015, http://www.defenseone.com/threats/2015/12/pentagon-nervous-about-russian-and-chinese-killer-robots/124465/?oref=d-river.

Tucker, Patrick. The Military Wants Smarter Insect Spy Drones." *Defense One*, December 23, 2014, http://www.defenseone.com/technology/2014/12/military-wants-smarter-insect-spy-drones/101970/

UNESCO Convention for the Protection of Cultural Property in the Event of Armed Conflict, May 14, 1954, 249 U.N.T.S. 240.

U.S. Opening Statement at the CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems, April 13, 2015, https://geneva.usmission.gov/2015/04/15/u-s-opening-statement-at-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapons-systems/

Velez-Green, Alexander. "The South Korean Sentry: A "Killer Robot" to Prevent War." *Lawfare Blog*, March 1, 2015.

Velozo, Thiago, and Lucas Bento. "ISIS Is Destroying Priceless Artifacts. Here's How to Stop Them." *The Diplomat*, March 17, 2015.

Wagner, Markus. "The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems." Vanderbilt Journal of Transnational Law 47(1) (2014): 33.

Wallach, Wendell. *A Dangerous Master*. New York: Basic Books, 2015.

Wallach, Wendell and Collin Allen. *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press, 2010.

Walsh, Fergus. "Robot operates inside eye in world first." *BBC News*, September 9, 2016, http://www.bbc.com/news/health-37246995

Webb, Jonathan. "Pneumatic octopus is first soft, solo robot." *BBC News*, August 25, 2016, http://www.bbc.com/news/science-environment-37169109.

Winfield, Alan. "Robotics: A Very Short introduction." Oxford: Oxford University Press, 2012.

Wittes, Benjamin, and Gabriella Blum. *The Future of Violence*. New York: Basic Books, 2015.

Witten, Ian H., and Eibe Frank. *Data Mining: Practical Machine Learning and Techniques with Java Implementations*. San Diego, CA: Morgan Kaufmann, 2006.

Work, Bob. "The Third Offset Strategy and its Implications for Partners and Allies." Deputy Secretary of Defense Speech, U.S. Department of Defense, January 28, 2015), http://www.defense.gov/News/Speeches/Speech-View/Article/606641/the-third-us-offset-strategy-and-its-implications-for-partners-and-allies.